Trabajo Monográfico

Probabilidades en Espacios Abstractos: una introducción.

Matías Carrasco

Orientador: Dr. Ricardo Fraiman

Licenciatura en Matemática
Facultad de Ciendicas
Universidad de la República
Uruguay

26 de Diciembre de 2005.

Resumen

Se consideran elementos aleatorios (e.a.) en un espacio de Banach separable $(X, \|\cdot\|)$. Se define el valor esperado de un e.a. en X por medio de la integral de Pettis. Se prueba una ley fuerte de los grandes números (l.f.g.n.) para una sucesión de e.a. $\{x_n\}$ independientes con idéntica distribución en X bajo la hipótesis de $\mathrm{E}\,\|x_1\| < \infty$. En el caso en que X es un espacio de Hilbert, se extienden otros resultado sobre la l.f.g.n. conocidos en el caso real, y se muestran contraejemplos de los mismo para espacios más generales.

En la segunda parte del trabajo consideramos X un espacio de Hilbert separable. Mediante la función característica (f.c.) de una medida de probabilidad en X se define un e.a. gaussiano. Se prueba la existencia utilizando el teorema de Minlos-Sazonov cuya prueba también damos aquí. Se demuestra un criterio para la compacidad relativa (de hecho un criterio para la tensión) debido a Prohorov y se prueba un teorema central del límite para e.a. independientes e identicamente distribuídos con varianza finita.

Palabras claves: probabilidades en espacios de Banach, ley de los grandes números, función característica, elementos aleatorios gaussianos, teorema central del límite.

Abstract

Random elements (r.e.) in a separable Banach space $(X, \|\cdot\|)$ are considered. We defined the expected value of a r.e. in X by means of the Pettis integral. We prove a strong law of large numbers (s.l.l.n.) for a sequence of independent with the same distribution r.e. $\{x_n\}$ in X, under the hypothesis of $\mathrm{E}\,\|x_1\| < \infty$. In the case in that X is a Hilbert space, we extend other known results on s.l.l.n. of the real case, and we show some counter-examples of the same one for more general spaces.

In the second part of the work we considered the case in that X is a separable Hilbert space. By means of the characteristic function (c.f.) of a probability measure in X we defined a gaussian r.e.. The existence is proven using the theorem of Minlos-Sazonov whose proof also we give here. We demonstrated a criterion for the relative compactness (in fact is a criterion for the tension property) due to Prohorov and a central limit theorem for independent, with the same distribution and finite variances r.e. is proven.

Key words and phrases: probabilities on Banach spaces, law of large numbers, characteristic functional, gaussian random elements, central limit theorem.

Índice general

Introducción.

El estudio de la teoría de probabilidades en espacios abstractos ocupa hoy en día un importante lugar en dicha disciplina. La necesidad de estudio en esta subrama de la teoría de probabilidades se manifiesta hace ya varios años, teniendo sus primeros impulsos en los trabajos de M. Frechet(1948) y posteriormente E. Mourier(1953). Por esos años, la consideración de los procesos estocásticos como *elementos aleatorios* en un espacio de funciones -es decir, una variable aleatoria a valores en un espacio funcional- realizada por Doob(1947), Mann(1951), Prohorov(1956) -y la escuela Rusa en general-, ha inspirado entre otras cosas el estudio de los teoremas límites para elementos aleatorios abstractos. Muchas de las leyes de los grandes números fueron generalizadas a elementos aleatorios que toman valores en un espacio abstracto. Por ejemplo, Mourier en los años 1953 y 1956 generalizó la ley fuerte para elementos aleatorios independientes e identicamente distribuídos a valores en un espacio de Banach separable. Posteriormente, Beck(1963) generalizó la ley fuerte para variables independientes no necesariamente con la misma distribución pero a expensas de exigirle al espacio una condición de convexidad, condición que resulta equivalente a la validez de la ley fuerte de Kolmogorov.[1] De este modo se ve claramente un vínculo muy profundo entre análisis y probabilidad; en este caso relacionados con aspectos geométricos de los espacios de Banach. Por ser este uno de los puntos de contacto de mayor actividad entre estas dos grandes disciplinas de la matemática, se encuentra una riqueza enorme de métodos nuevos no provenientes de ninguna de ellas en particular. El estudio de la convergencia débil de medidas y los teoremas límites en general también tuvieron sus generalizaciones.

La importancia de estos estudios no es simplemente la importancia que tiene toda abstracción, que de hecho es mucha. No sólo se trata de expandir las fronteras de la misma, sino que las aplicaciones tanto de carácter teórico como práctico son relevantes. Las aplicaciones del estudio de la convergencia débil en espacios más abstractos y los teoremas límites en general

[1] Para elementos aleatorios independientes no necesariamente con la misma distribución.

tanto dentro de la propia disciplina como su vinculación con otras, ha sido muy fructífero. Otro aspecto importante son las aplicaciones estadísticas. En contraste con un enfoque que consiste en un tratamiento de datos infinito-dimensionales mediante aproximaciones con bases adecuadas, considerar la propia naturaleza infinito dimensional de los datos puede ser muy provechoso. Así, la necesidad de tener espacios y teorías capaces de manipular dichos objetos es muy importante; inclusive para problemas muy sencillos en estadística como pueden ser el problema de posición ó el problema de regresión. Por ejemplo, surge la necesidad de definir de forma adecuada el concepto de mediana, el concepto de forma cuadrática, entre muchos otros problemas estadísticos en espacios de Banach. Son temas que no hemos tratado aquí pero que reflejan un panorama bastante amplio de cosas por hacer, es decir, en forma un tanto exagerada podríamos afirmar que hay mucha estadística -infinito dimensional- que requiere ser estudiada.

El objetivo de esta monografia es el de exponer en forma detallada y autocontenida algunas partes importantes de la teoría. En el primer capítulo introducimos los conceptos básicos, las definiciones necesarias y los problemas que se plantean al extender los tipos de espacios con los que se trabaja. Se define el concepto de *valor esperado* por medio de la integral de Pettis para espacios normados y se prueba un teorema de existencia.

En el capítulo dos se trabaja con la ley de los grandes números. La teoría en espacios de Hilbert separables se desarrolla sin problemas utilizando prácticamente los mismos métodos que en el caso de variables reales. Así, son válidos los teoremas de Rademacher-Mensov para el caso ortogonal, las desigualdades de Kolmogorov y ambas leyes fuertes de Kolmogorov para el caso independiente. Sin embargo, para espacios de Banach la teoría es más trabajosa, valiendo sí el teorema de Mourier que mecionamos arriba pero siendo falsas en un contexto general las leyes fuertes y débiles para el caso independiente cuando se imponen condiciones en las varianzas. Hecho fundamental es la invalidez de la propiedad de aditividad de la varianza para elementos independientes. La generalización de estos teoremas en este caso requiere la exigencia de condiciones de convexidad como ya dijimos.

En el capitulo tres estudiamos la convergencia débil y el teorema central del límite. La primera en un contexto bastante general y el segundo exclusivamente para el caso de elementos aleatorios i.i.d. a valores en un espacio de Hilbert. Se estudia el teorema de Prohorov donde se afirma que la propiedad de *tensión* en una familia de probabilidades es suficiente para la compacidad relativa y que es también necesaria en el caso separable y completo. Estudiamos también un criterio para la tensión debido a Prohorov para medidas en espacios de Hilbert. El método utilizado es exclusivamente el de las *funciones características* o transformada de

Fourier, lo cual implica el estudio de los *S-operadores*. Se define así la distribución gaussiana en términos de su función característica mediante el teorema de Milnos-Sazonov, que es una extensión del teorema de Bochner para el caso infinito dimensional.

Por último mencionamos que hemos estudiado una pequeña parte de la teoría. Muchos resultados existen; como por ejemplo teoremas límites para arreglos triangulares que cumplan la condición de Lindeberg pueden ser extendidos a espacios de Hilbert. También el estudio de las distribuciones infinitamente divisibles en estructuras algebraicas más generales. En este sentido este trabajo consiste en una “caja de herramientas”, tanto para proseguir en estudios de carácter teórico como estudios aplicados.

Capítulo 1

Elementos Aleatorios en Espacios Métricos.

1.1. Definición de elemento aleatorio.

Sea (M, d) un espacio métrico, denotamos por $\mathcal{B}_M$ la σ-álgebra de Borel en M -la menor σ-álgebra de M que contiene a los subconjuntos abiertos de M-. En este contexto decimos que una aplicación $T : M \to N$, donde N es otro espacio métrico, es *Borel medible* si $T^{-1}(B) = \{x \in M : T(x) \in B\} \in \mathcal{B}_M$ para todo subconjunto boreliano $B \in \mathcal{B}_N$. En lo que sigue trabajaremos siempre con un espacio de probabilidad fijo $(\Omega, \mathcal{A}, \mathbf{P})$.

Definición 1.1. *Una aplicación $v : \Omega \to M$ se dice un elemento aleatorio (e.a.) en M, si $v^{-1}(B) = \{\omega \in \Omega : v(\omega) \in B\} \in \mathcal{A}$ para todo subconjunto $B \in \mathcal{B}_M$; es decir, la aplicación v es $(\mathcal{A}, \mathcal{B}_M)$-medible.*

Si $M = \mathbb{R}$ con la métrica usual, la noción de elemento aleatorio es la misma que la de variable aleatoria. Lo mismo ocurre en $\mathbb{R}^n$, pues los abiertos son uniones numerables de "rectángulos" abiertos $I_1 \times I_2 \times \cdots \times I_n$, donde cada $I_i = (a_i, b_i)$ es un intervalo abierto de $\mathbb{R}$. Luego una función $\vec{x} = (x_1, \cdots, x_n) : \Omega \to \mathbb{R}^n$ es un elemento aleatorio en $\mathbb{R}^n$ si y sólo si cada $x_i : \Omega \to \mathbb{R}$ es una variable aleatoria.

Obtenemos otro ejemplo si consideramos $\mathbf{s}$ el conjunto de las sucesiones en $\mathbb{R}$ con la métrica producto: si $x = \{x_n\}$ e $y = \{y_n\}$ entonces

$$d(x, y) = \sum_{n=1}^{+\infty} \frac{1}{2^n} \frac{|x_n - y_n|}{1 + |x_n - y_n|}$$

Las proyecciones $\pi_n : \mathbf{s} \to \mathbb{R}$ dadas por $\pi_n(x) = x_n$ son continuas y por ende Borel medibles.

Sean $x \in \mathbf{s}$, $r > 0$ y consideremos

$$B(x,r) = \{y \in \mathbf{s} : d(y,x) < r\}$$

la bola abierta de centro x y radio r. Para cada entero positivo m y cada $\epsilon > 0$ definimos los conjuntos abiertos

$$N^x_{m,\epsilon} = \left\{ y : \frac{1}{2^i}\frac{|y_i - x_i|}{1 + |y_i - x_i|} < \epsilon \text{ para } i = 1, 2, \dots, m \right\}$$

Sea $m \geq 1$ tal que $\sum_{n \geq m+1} 2^{-n} < r/2$. Entonces si $y \in N^x_{m,\epsilon}$ para $\epsilon = r/2m$, tenemos que $d(y,x) < r$ de donde el conjunto abierto $N^x_{m,\epsilon} \subseteq B(x,r)$. Luego la familia de estos conjuntos forman una base para la topología de $\mathbf{s}$, y más aún, todo abierto de $\mathbf{s}$ es unión numerable de conjuntos de esta familia; es decir, $\mathbf{s}$ es separable.

Sea $v : \Omega \to \mathbf{s}$ un elemento aleatorio en $\mathbf{s}$. Escribimos para cada $\omega \in \Omega$,

$$v(\omega) = (v_1(\omega), v_2(\omega), \dots)$$

donde cada $v_i : \Omega \to \mathbb{R}$. Luego como $v_i = \pi_i(v)$, tenemos que v_i es una variable aleatoria para cada $i \geq 1$.

Reciprocamente, supongamos que cada v_i es una variable aleatoria. Como

$$(*) \quad v^{-1}(I_1 \times \cdots \times I_m \times \mathbb{R} \times \cdots) = \bigcap_{i=1}^{m} v_i^{-1}(I_i)$$

tenemos que si cada I_i es un abierto en $\mathbb{R}$, $(*)$ pertenece a $\mathcal{A}$, por lo que v es un elemento aleatorio en $\mathbf{s}$. En resumen, $v = (v_1, v_2, \cdots)$ es un elemento aleatorio en $\mathbf{s}$ si y sólo si cada v_i es una variable aleatoria en $\mathbb{R}$. Luego podemos pensar a los elementos aleatorios en $\mathbf{s}$ como una sucesión de variables aleatorias en $\mathbb{R}$.

Es importante notar que en los tres ejemplos que vimos, pudimos caracterizar a los elementos aleatorios en términos de sus coordenadas utilizando fuertemente que dichos espacios eran espacios métricos separables. Ésta es una característica que como veremos más adelante se mantiene en los espacios separables en general.

Muchas veces se utiliza la noción de elemento aleatorio de *Radon*, que exige además que la distribución de dicho elemento aleatorio sea una medida de probabilidad *tensa* en M[1]. Una noción ligeramente más débil que la definición que nosotros dimos es la siguiente: si el

[1]Estos conceptos los definiremos más adelante.

espacio M es un espacio de Banach y B^* denota su dual topológico, se dice que v es un elemento aleatorio -en el sentido débil- si para todo $f \in B^*$, $f(v)$ es una variable aleatoria. Esta última definición es más tratable si el espacio B cumple la siguiente propiedad: existe un conjunto numerable D, contenido en la bola unidad B^*, tal que para todo $x \in B$

$$\|x\| = \sup_D f(x)$$

pues hace $\|\cdot\|$ una variable aleatoria. No entraremos en estos temas, pero mencionamos que en el caso separable toda noción razonable de medibilidad de los elementos aleatorios coincide con la que nosotros dimos.

1.2. Propiedades básicas de los elementos aleatorios.

En esta sección trataremos de extender las propiedades básicas de las variables aleatorias en $\mathbb{R}$ a elementos aleatorios en un espacio métrico general.

Proposición 1.1. *Sean M y M_1 espacios métricos y $T : M \to M_1$ una función Borel medible. Entonces si v es un e.a. en M, $v_1 = T(v)$ es un e.a. en M_1.*

Demostración. Basta observar que la composición de una función $(\mathcal{A}, \mathcal{B}_M)$-medible con una aplicación Borel medible, es también $(\mathcal{A}, \mathcal{B}_M)$-medible. □

Proposición 1.2. *Sea $\{E_n\}_{n\geq 1}$ una sucesión de conjuntos en $\mathcal{A}$ tales que $E_n \cap E_m = \emptyset$ si $n \neq m$, y $\Omega = \bigcup_{n\geq 1} E_n$. Sea además $\{x_n\}$ una sucesión en M. Entonces la función $v : \Omega \to M$ tal que $v(\omega) = x_n$ cuando $\omega \in E_n$, es un elemento aleatorio en M.*

Demostración. Sea $B \in \mathcal{B}_M$. Entonces si $\{x_{n_k}\}_{k\geq 1}$ es el conjunto de valores de v tales que $x_{n_k} \in B$, tenemos que $v^{-1}(B) = \bigcup_{k\geq 1} E_{n_k} \in \mathcal{A}$. □

Proposición 1.3. *Sea $\{v_n\}_{n\geq 1}$ una sucesión de e.a. en M tal que $v_n(\omega) \to v(\omega)$ para cada $\omega \in \Omega$. Entonces la aplicación definida por $\omega \mapsto v(\omega)$ es un e.a. en M.*

Demostración. Necesitamos probar que v es Borel medible. Como podemos generar $\mathcal{B}_M$ con los conjuntos cerrados, bastará probar que $v^{-1}(C) \in \mathcal{A}$ para todo $C \subseteq M$ cerrado. La clave consiste en escribir $v^{-1}(C)$ de una forma conveniente.

Sea $v(\omega) \in C$. Como $v_n(\omega) \to v(\omega)$, dado $k \geq 1$

$$d(v_n(\omega), v(\omega)) < \frac{1}{k}$$

a partir de un m -que depende de ω-. Entonces

$$v_n(\omega) \in C_k = \{x \in M : d(x, C) < \frac{1}{k}\}$$

para todo $n \geq m$.[2] Es decir que $\omega \in \bigcap_{n\geq m} v_n^{-1}(C_k)$. Como esto debe ocurrir para todo $k \geq 1$ concluimos que $\omega \in \bigcap_{k\geq 1} \bigcup_{m\geq 1} \bigcap_{n\geq m} v_n^{-1}(C_k)$.

Reciprocamente si $\omega \in \bigcap_{k\geq 1} \bigcup_{m\geq 1} \bigcap_{n\geq m} v_n^{-1}(C_k)$, entonces dado $k \geq 1$, existe $m \geq 1$ tal que $\omega \in \bigcap_{n\geq m} v_n^{-1}(C_k)$. Luego para cada $n \geq m$, existe $x_n \in C$ tal que

$$d(x_n, v_n(\omega)) < \frac{1}{k}$$

Como $v_n(\omega) \to v(\omega)$, si $n \geq n_0$ tenemos que $d(v_n(\omega), v(\omega)) < \frac{1}{k}$, de donde

$$d(v(\omega), x_n) \leq d(v(\omega), v_n(\omega)) + d(v_n(\omega), x_n) < \frac{2}{k}$$

si $n \geq \max\{m, n_0\}$. Esto implica que $d(v(\omega), C) = 0$, es decir $v(\omega) \in \overline{C} = C$ pues C es cerrado. Es decir, $\omega \in v^{-1}(C)$. En resumen,

$$v^{-1}(C) = \bigcap_{k\geq 1} \bigcup_{m\geq 1} \bigcap_{n\geq m} v_n^{-1}(C_k)$$

que es un conjunto medible -de $\mathcal{A}$-, pues cada C_k es abierto en M y cada v_n es un elemento aleatorio en M. $\square$

Usaremos estas dos proposiciones para probar que en un espacio métrico separable M, todo elemento aleatorio es el límite uniforme de una sucesión de elementos aleatorios "discretos" que toman una cantidad numerable de valores. Para esto precisamos el siguiente lema.

Lema 1.1. *Sea M un espacio métrico separable. Entonces dado $\lambda > 0$ arbitrario, existe una función $T : M \to M$ Borel medible, tal que T toma un conjunto numerable de valores y $d(T(x), x) < \lambda$ para todo $x \in M$.*

Demostración. Sea $\{x_n\}$ un subconjunto denso numerable en M. Consideremos los conjuntos $E_1 = B(x_1, \lambda)$ y

$$E_n = B(x_n, \lambda) - \bigcup_{i=1}^{n-1} B(x_i, \lambda) \text{ para } n > 1$$

[2] Aquí $d(x, C) = \inf\{d(x, y) : y \in C\}$.

Entonces los conjuntos $\{E_n\}_{n\geq 1}$ son disjuntos dos a dos y cubren M. Podemos definir entonces $T(x) = x_n$ si $x \in E_n$. Es claro que T es Borel medible, pues si $\{x_{n_k}\}_{k\geq 1}$ son aquellos elementos del conjunto denso que están en B, para un $B \in \mathcal{B}_M$ dado, entonces

$$T^{-1}(B) = \bigcup_{k\geq 1} E_{n_k} \in \mathcal{B}_M$$

ya que cada E_n es un boreliano. Además dado $x \in M$, $x \in E_n$ para algún n, por lo que $T(x) = x_n$ y $d(T(x), x) = d(x_n, x) < \lambda$. $\square$

Proposición 1.4. *Sea M un espacio métrico separable. Un mapa $v : \Omega \to M$ es un e.a. en M, si y sólo si existe una sucesión $\{v_n\}$ de e.a. discretos en M que converge uniformemente a v.*

Demostración. Supongamos que $v : \Omega \to M$ es un elemento aleatorio en M. Para cada $n \geq 1$, como M es separable sabemos que existe $T_n : M \to M$ Borel medible, que toma una cantidad numerable de valores y tal que $d(T_n(x), x) < n^{-1}$ para todo $x \in M$. Entonces definimos $v_n : \Omega \to M$ dado por $v_n = T_n(v)$. Luego v_n es un e.a. discreto en M tal que

$$\sup_{\omega\in\Omega} d(v_n(\omega), v(\omega)) < \frac{1}{n}$$

de donde v_n converge uniformemente en Ω a v.

Reciprocamente supongamos que existe tal sucesión $\{v_n\}$ de e.a. discretos que converge uniformemente a v. Como en particular $v_n(\omega) \to v(\omega)$ para todo $\omega \in \Omega$, sabemos por la proposición anterior que v es un elemento aleatorio en M. $\square$

Dados dos elementos aleatorios u y v en M la función $x = d(u, v)$ -en un contexto general- no tiene por qué en principio ser medible, es decir, no tiene por qué ser una variable aleatoria. Como veremos más adelante, muchos resultados y definiciones se expresan en términos de la métrica d, lo cual restringe en una primera instancia la clase de espacios que nos interesan. Una condición suficiente para que x sea una variable aleatoria es que el espacio M sea separable. Denotamos $\mathcal{B}_{M_1} \times \mathcal{B}_{M_2}$ la σ-álgebra generada por los conjuntos de la forma $A_1 \times A_2$ tales que $A_i \in \mathcal{B}_{M_i}$.

Proposición 1.5. *Sean u y v e.a. en un espacio métrico M. Si M es separable, entonces $d(u, v)$ es una variable aleatoria.*

Demostración. Sabemos que la función $d : M \times M \to \mathbb{R}$ es continua, luego es Borel medible. También sabemos que $u, v : \Omega \to M$ son funciones $(\mathcal{A}, \mathcal{B}_M)$-medibles, si y sólo si el mapa

$\omega \to (u(\omega), v(\omega))$ es medible sobre la σ-álgebra producto $\mathcal{B}_M \times \mathcal{B}_M$ en $M \times M$. Como M es separable sabemos que $\mathcal{B}_{M\times M} = \mathcal{B}_M \times \mathcal{B}_M$. Lo anterior implica que al ser u y v medibles, (u, v) sea un e.a. en $M \times M$. Luego la composición $x = d(u, v)$ es un e.a. en $\mathbb{R}$, es decir, una variable aleatoria. □

Consideremos un espacio métrico M discreto con cardinal superior al de $\mathbb{R}$. Como todo subconjunto de M es abierto,

$$\mathcal{B}_M \times \mathcal{B}_M = \sigma(\{A_1 \times A_2 : A_i \subseteq M_i \text{ abierto}\}) = \sigma(\{A_1 \times A_2 : A_i \subseteq M_i\})$$

Además $\Delta = \{(x, y) : x = y\}$ es un subconjunto cerrado de $M \times M$[3] y por ende $\Delta \in \mathcal{B}_{M\times M}$. En realidad $\mathcal{B}_{M\times M} = \mathcal{P}(M \times M)$ pues el producto de discretos es también discreto. Pero $\Delta \notin \mathcal{B}_M \times \mathcal{B}_M$. (Ver Halmos, Measure Theory, Problema 2 pág. 261).

Consideremos $\Omega = M\times M$ y $\mathcal{A} = \mathcal{B}_M\times\mathcal{B}_M$. Entonces las proyecciones $u, v : \Omega \to M$ dadas por $u(x, y) = x$ y $v(x, y) = y$ son medibles como funciones entre estos espacios. Veamos que $x = d(u, v)$ no es una variable aleatoria. Lo que ocurre es que Δ es un boreliano en $M \times M$, pero $\Delta \notin \mathcal{A}$ por lo que (u, v) no es un elemento aleatorio en $M\times M$. Así, como $\Delta = d^{-1}(\{0\})$ y $\{0\} \in \mathcal{B}_{\mathbb{R}}$, $x^{-1}(\{0\}) = \Delta \notin \mathcal{A}$ con lo cual x no es una variable aleatoria.

Debido a esto nos restringiremos a espacios métricos (M, d) separables para dar las siguientes definiciones. Igualmente siempre que sea posible dar los resultados en un contexto más general, lo haremos.

Definición 1.2. *Sea $\{v_n\}_{n\geq 1}$ una sucesión de e.a. y v un e.a. en un espacio métrico separable M. Decimos que v_n converge a v:*

- *con probabilidad 1 o casi seguramente ($v_n \to v$ c.s.) si*

$$P\left(\{\omega \in \Omega : d(v_n(\omega), v(\omega)) \to 0\}\right) = 1$$

- *en probabilidad ($v_n \to^P v$) si para todo $\epsilon > 0$,*

$$P\left(\{\omega \in \Omega : d(v_n(\omega), v(\omega)) \geq \epsilon\}\right) \to 0$$

- *en la r-media ($v_n \to^r v$) con $r > 0$, si*

$$\mathrm{E}d(v_n, v)^r \to 0$$

[3] Esto es pues $\Delta = d^{-1}(\{0\})$.

Es decir, decimos que v_n converge a v casi seguramente, en probabilidad o en r-media, cuando la variable aleatoria $x_n = d(v_n, v)$ converge a cero casi seguramente, en probabilidad o en r-media, respectivamente.

Como nuestro primer objetivo es la ley de los grandes números, no damos aquí la definición de convergencia en distribución que veremos más adelante. Cuando hablemos de la ley fuerte de los grandes números, nos estaremos refiriendo a la convergencia de los promedios en el sentido casi seguro, y si usamos la palabra débil, será en el sentido de convergencia en probabilidad.

Las relaciones entre los distintos tipos de convergencia son exactamente las mismas que en el caso real. Es decir, por ejemplo la convergencia c.s. implica la convergencia en probabilidad. Otra propiedad que extendemos es por ejemplo la desigualdad de Markov. Como $d(u, v)$ es una variable aleatoria no negativa, si existe $\mathrm{E}d(u, v)^r$ para algún $r > 0$, tenemos que

$$P\left(d(u, v) \geq \epsilon\right) \leq \frac{1}{\epsilon^r}\mathrm{E}d(u, v)^r$$

para todo $\epsilon > 0$. Luego si $v_n \to^r v$ para algún $r > 0$, aplicando esta desigualdad vemos que $v_n \to^P v$.

Sea $\{v_n\}$ una sucesión de e.a. y v un e.a. en un espacio métrico separable M. Si existe un $r > 0$ tal que

$$\sum_{n=1}^{+\infty} \mathrm{E}d(v_n, v)^r < +\infty$$

entonces $v_n \to v$ c.s.. Las pruebas detalladas de estos resultados no las hacemos aquí pues son totalmente análogas a las mismas para el caso real. De todos modos se puede consultar Padgett-Taylor[3].

1.3. Elementos aleatorios en espacios normados.

En esta sección veremos algunas propiedades de los elementos aleatorios cuando el espacio métrico M es un espacio normado, es decir, un espacio vectorial E con una norma que denotaremos por $\|\cdot\|$. En estas condiciones tenemos definido el espacio dual E^* de E, formado por los funcionales lineales $f : E \to \mathbb{R}$ continuos, es decir aquellos para los cuales

$$\|f\| = \sup_{\|x\|=1} |f(x)| < +\infty$$

Éste también es un espacio normado -en realidad es de Banach- con esta misma norma.

Como la norma es una función continua tenemos que si $v : \Omega \to E$ es un elemento aleatorio en E, entonces $\|v\|$ es una variable aleatoria[4]. Del mismo modo cada funcional $f \in E^*$ es una función continua, por lo tanto $f(v)$ es una variable aleatoria para cada $f \in E^*$. En el caso separable el hecho de ser una variable aleatoria $f(v)$ para cada $f \in E^*$ nos garantiza que v sea un elemento aleatorio. La idea de la demostración de esto la utilizaremos bastante, por lo que nos será mejor disponer del siguiente lema.

Lema 1.2. *Sea* $(E, \|\cdot\|)$ *un espacio normado separable. Si* $\{x_n\}$ *es un subconjunto denso en* $S(0,1) = \{x \in E : \|x\| = 1\}$, *sea* $g_n \in E^*$ *tal que* $\|g_n\| = 1$ *y* $g_n(x_n) = 1$ *para cada* $n \geq 1$. *Entonces*[5]

$$B[0,1] = \bigcap_{n\geq 1} \{x \in E : g_n(x) \leq 1\}$$

Demostración. Si $x \in B[0,1]$, entonces para cada $n \geq 1$ tenemos que $|g_n(x)| \leq \|g_n\| \, \|x\| = \|x\| \leq 1$, de donde $x \in A = \bigcap_{n\geq 1}\{x \in E : g_n(x) \leq 1\}$.

Sea $x \notin B[0,1]$, es decir $\|x\| > 1$. Entonces por la densidad de $\{x_n\}$, existe $\{x_{n_k}\}$ tal que $\left\| \frac{x}{\|x\|} - x_{n_k} \right\| < \frac{1}{k}$. Luego tenemos que

$$\left| g_{n_k}\left(\frac{x}{\|x\|} - x_{n_k} \right) \right| \leq \left\| \frac{x}{\|x\|} - x_{n_k} \right\| < \frac{1}{k}$$

Pero como $g_{n_k}\left(\frac{x}{\|x\|} - x_{n_k}\right) = \frac{g_{n_k}(x)}{\|x\|} - 1$, tenemos que $\frac{|g_{n_k}(x) - \|x\||}{\|x\|} \to_k 0$, es decir

$$g_{n_k}(x) \to_k \|x\|$$

Luego existe k suficientemente grande tal que $g_{n_k}(x) > 1$. Entonces $x \notin A$. □

Proposición 1.6. *Sea* $(E, \|\cdot\|)$ *un espacio normado separable. Consideremos la familia* $\mathcal{F}$ *de los conjuntos*

$$\{x \in E : (f_1(x), \ldots, f_n(x)) \in B\}$$

dende $n \geq 1$, $f_i \in E^*$ *y* $B \in \mathcal{B}_{\mathbb{R}^n}$ *para* $i = 1, \ldots, n$. *Entonces* $\mathcal{B}_E = \sigma(\mathcal{F})$.

Demostración. Consideremos la familia ligeramente distinta $\mathcal{F}_0$ formada por los conjuntos de la forma

$$\{x \in E : f_1(x) \in B_1, \ldots, f_n(x) \in B_n\}$$

[4]Observar que no es necesario que E sea separable.

[5]Denotamos a la bola cerrada de esta manera

con $n \geq 1$, $f_i \in E^*$ y $B_i \in \mathcal{B}_{\mathbb{R}}$. Como $\mathcal{F}_0 \subseteq \mathcal{F}$, tenemos que $\sigma(\mathcal{F}_0) \subseteq \sigma(\mathcal{F})$ y como los funcionales son continuos es claro que $\sigma(\mathcal{F}_0) \subset \mathcal{B}_E$. Luego bastará probar que $\mathcal{B}_E \subseteq \sigma(\mathcal{F}_0)$. Pero al ser E separable todo abierto es unión numerable de bolas abiertas, las cuales son uniones numerables de bolas cerradas, por lo que es suficiente ver que estas últimas pertenecen a $\sigma(\mathcal{F}_0)$.

Sea $B[x_0, r]$ una bola cerrada de centro x_0 y radio $r > 0$ en E. Definimos $T : E \to E$ dada por $T(x) = \frac{1}{r}(x - x_0)$, entonces T es un homeomorfismo y por el lema anterior tenemos que

$$B[x_0, r] = T^{-1}(B[0,1]) = T^{-1}(\bigcap_{n\geq 1}\{x \in E : g_n(x) \leq 1\}) =$$

$$\bigcap_{n\geq 1} T^{-1}(\{x \in E : g_n(x) \leq 1\}) = \bigcap_{n\geq 1}\{x : g_n(T(x)) \leq 1\} =$$

$$\bigcap_{n\geq 1}\{x : g_n(x) \leq r + g_n(x_0)\} = \bigcap_{n\geq 1} g_n^{-1}((-\infty, \alpha_n]) \in \sigma(\mathcal{F}_0)$$

□

Más aún, hemos probado que $\mathcal{B}_E = \sigma(\{f^{-1}((-\infty, r]) : r \in \mathbb{R}, \quad f \in E^*\})$.

Corolario 1.1. *Sea $(E, \|\cdot\|)$ un espacio normado separable. Si $v : \Omega \to E$ es tal que $f(v)$ es una variable aleatoria para cada $f \in E^*$, entonces v es un e.a. en E.*

Demostración. Como $f(v) : \Omega \to \mathbb{R}$ es medible el conjunto $\{f(v) \in B\}$ es medible para todo $f \in E^*$ y todo $B \in \mathcal{B}_{\mathbb{R}}$. Es decir, el conjunto

$$\{v \in f^{-1}(B)\} \in \mathcal{A}$$

para todo $f \in E^*$ y $B \in \mathcal{B}_{\mathbb{R}}$. Entonces v es $(\mathcal{A}, \sigma(\mathcal{F}_0))$-medible y como por el lema $\sigma(\mathcal{F}_0) = \mathcal{B}_E$, v es un e.a. en E. □

Corolario 1.2. *Sean $v_1, \ldots, v_n$ elementos aleatorios en un espacio normado separable E. Si $\lambda_1, \ldots, \lambda_n$ son reales arbitrarios, entonces $\sum_{i=1}^n \lambda_i v_i$ es un elemento aleatorio en E.*

Demostración. Sea $f \in E^*$ arbitrario. Entonces $f\left(\sum_{i=1}^n \lambda_i v_i\right) = \sum_i^n \lambda_i f(v_i)$, y como una combinación lineal de variables aleatorias -en $\mathbb{R}$- es también una variable aleatoria, concluimos por el corolario anterior que $\sum_{i=1}^n \lambda_i v_i$ es un elemento aleatorio en E. □

Como ejemplo podemos caracterizar utilizando lo que tenemos hasta ahora, cuáles son los elementos aleatorios en ciertos espacios. Si denotamos l^1 el espacio vectorial de las sucesiones reales absolutamente sumables con la norma $\|x\| = \sum_{n\geq 1} |x_n|$, entonces $v = (v_1, v_2, \cdots)$ es un elemento aleatorio en l^1, si y sólo si $\{v_i\}$ es una sucesión de variables aleatorias absolutamente sumables. Si v es un elemento aleatorio en l^1, entonces $\pi_i \circ v = v_i$ es una variable aleatoria pues π_i -la proyección canónica- es una función continua. Reciprocamente, supongamos que $\{v_n\} \in l^1$ es una sucesión de variables aleatorias. Sea $f \in (l^1)^*$, entonces existe $\{a_n\} \in l^\infty$ -las sucesiónes reales acotadas con la norma del supremo- tal que $f(x) = \sum_{n\geq 1} a_n x_n$ para todo $x \in l^1$. Luego

$$f(v)(\omega) = \sum_{n\geq 1} a_n v_n(\omega) = \lim_{N\to+\infty} \sum_{n=1}^{N} a_n v_n(\omega)$$

por lo que es medible al ser límite de funciones medibles. Como f es arbitrario, tenemos que v es un elemento aleatorio en l^1.

Estos mismos argumentos nos permiten afirmar que si $l^p = \{x = \{x_n\} : \sum_{n\geq 1} |x_n|^p < +\infty\}$ con la norma $\|x\| = \left(\sum_{n\geq 1} |x_n|^p\right)^{\frac{1}{p}}$ para $p > 1$, entonces los elementos aleatorios en l^p son las sucesiones $\{v_n\}$ de variables aleatorias en l^p.

Proposición 1.7. *Sean E un espacio normado*[6]*, v un elemento aleatorio en E y α una variable aleatoria. Entonces αv es un elemento aleatorio en E.*

Demostración. Supongamos primero que α es discreta, es decir toma una cantidad numerable de valores; a saber $\{x_1, x_2, \ldots\}$. Sea $B \in \mathcal{B}_E$, entonces

$$(\alpha v)^{-1}(B) = (\alpha v)^{-1}(B) \cap \bigcup_{n\geq 1} E_n = \bigcup_{n\geq 1} (\alpha v)^{-1}(B) \cap E_n$$

donde $E_n = \{\omega \in \Omega : \alpha(\omega) = x_n\}$. Como $(\alpha v)^{-1}(B) = \{\omega : \alpha(\omega)v(\omega) \in B\}$ y en E_n $\alpha(\omega) = x_n$, tenemos que

$$(\alpha v)^{-1}(B) \cap E_n = \{\omega : x_n v(\omega) \in B\} = \{\omega : v(\omega) \in T_n^{-1}(B)\}$$

con $T_n : E \to E$ dada por $T_n(v) = x_n v$. Como T_n es continua, al ser B un boreliano también lo es su preimagen por T_n. Luego como v es un elemento aleatorio, $(\alpha v)^{-1}(B) \cap E_n \in \mathcal{A}$, de donde también $(\alpha v)^{-1}(B) \in \mathcal{A}$.

En el caso general, sabemos que existe $\{\alpha_n\}$ una sucesión de variables aleatorias discretas que convergen puntualmente a α. Nuevamente por la continuidad de la multiplicación por escalares, tenemos que

$$\alpha_n(\omega)v(\omega) \to_n \alpha(\omega)v(\omega)$$

[6]No necesariamente separable.

para cada $\omega \in \Omega$, y por lo tanto αv es un elemento aleatorio por ser límite de elementos aleatorios en E. $\square$

Si v es un elemento aleatorio en un espacio métrico M, entonces éste define una medida de probabilidad en $(M, \mathcal{B}_M)$ dada por

$$P_v(B) = P\left(v^{-1}(B)\right) = P(v \in B)$$

A esta medida de probabilidad P_v la llamaremos distribución o ley de v. Las siguientes definiciones que vamos a dar las podemos formular sin problemas en un espacio métrico en general, a pesar de estar esta sección destinada a estudiar elementos aleatorios en espacios normados.

Definición 1.3. *Sean v_1 y v_2 dos elementos aleatorios en un espacio métrico M. Decimos que v_1 y v_2 son identicamente distribuidos si*

$$P(v_1 \in B) = P(v_2 \in B)$$

para todo subconjunto boreliano $B \in \mathcal{B}_M$. Es decir, si las medidas de probabilidad inducidas por ambos en M coinciden. Diremos que una familia de elementos aleatorios en M es identicamente distribuida, si dos cualquiera de la familia son identicamente distribuidos.

Definición 1.4. *Dado un conjunto finito de elementos aleatorios $\{v_1, \ldots, v_n\}$ en un espacio métrico M, decimos que son independientes si*

$$P(v_1 \in B_1, \ldots, v_n \in B_n) = P(v_1 \in B_1) \cdots P(v_n \in B_n)$$

para toda colección $B_1, \ldots, B_n \in \mathcal{B}_M$. Decimos que una colección arbitraria de elementos aleatorios en M es independiente si todo subconjunto finito de ésta, está formado por elementos aleatorios independientes.

Proposición 1.8. *Sean v_1 y v_2 elementos aleatorios independientes e identicamente distribuidos en un espacio métrico M, y sea ϕ una función Borel medible de M en otro espacio métrico M_1. Entonces $\phi(v_1)$ y $\phi(v_2)$ son elementos aleatorios independientes e identicamente distribuidos en M_1.*

Demostración. Sea $B_1 \in \mathcal{B}_{M_1}$, entonces $B = \phi^{-1}(B_1) \in \mathcal{B}_M$, de donde

$$P(\phi(v_1) \in B_1) = P\left(v_1 \in \phi^{-1}(B_1)\right) = P(v_1 \in B)$$

$$= P(v_2 \in B) = P\left(v_2 \in \phi^{-1}(B_1)\right) = P(\phi(v_2) \in B_1)$$

por lo que $\phi(v_1)$ y $\phi(v_2)$ son identicamente distribuidos.

Sean $B_1, B_2 \in \mathcal{B}_{M_1}$, entonces

$$P(\phi(v_1) \in B_1, \phi(v_2) \in B_2) = P\left(v_1 \in \phi^{-1}(B_1), v_2 \in \phi^{-1}(B_2)\right)$$

$$= P\left(v_1 \in \phi^{-1}(B_1)\right) P\left(v_2 \in \phi^{-1}(v_2)\right) = P(\phi(v_1) \in B_1) P(\phi(v_2) \in B_2)$$

y tenemos así la independencia. □

Es claro que este resultado se extiende con la misma demostración para una sucesión de elementos aleatorios.

Nuestro próximo objetivo es dar algunas condiciones bajo las cuales podamos garantizar que dos elementos aleatorios en un espacio normado sean identicamente distribuidos.

Definición 1.5. *Sea $\mathcal{F}$ una familia de subconjuntos en $\mathcal{B}_M$. Decimos que $\mathcal{F}$ es una clase determinante para $\mathcal{B}_M$ si dadas dos medidas de probabilidad en $\mathcal{B}_M$, P y Q, tales que $P(A) = Q(A)$ para todo $A \in \mathcal{F}$; necesariamente debe ser $P(B) = Q(B)$ para todo $B \in \mathcal{B}_M$.*

Proposición 1.9. *Sean E un espacio normado separable y $v_1, v_2 : \Omega \to E$ dos elementos aleatorios en E. Entonces, v_1 y v_2 son identicamente distribuidos si y sólo si $f(v_1)$ y $f(v_2)$ son variables aleatorias con igual distribución para todo $f \in E^*$.*

Demostración. Si v_1 y v_2 son identicamente distribuidos, como $f \in E^*$ es Borel medible por la proposición anterior $f(v_1)$ y $f(v_2)$ también lo son.

Supongamos que $f(v_1)$ tiene la misma distribución que $f(v_2)$ para todo $f \in E^*$. Esto quiere decir que $P(f(v_1) \in B) = P(f(v_2) \in B)$ para todo $f \in E^*$ y $B \in \mathcal{B}_{\mathbb{R}}$.

Como antes sea $\mathcal{F}$ la familia de subconjuntos de $\mathcal{B}_E$ formada por los conjuntos de la forma

$$\{x \in E : (f_1(x), \cdots, f_n(x)) \in B\}$$

con $f_i \in E^*$ y $B \in \mathcal{B}_{\mathbb{R}^n}$. Es claro que $\mathcal{F}$ es un álgebra de conjuntos, por lo que bastará probar que las medidas de probabilidad inducidas por v_1 y v_2 coinciden en $\mathcal{F}$.

Para esto, sea $A \in \mathcal{F}$ con

$$A = \{x : (f_1(x), \cdots, f_n(x)) \in B\} \text{ donde } f_i \in E^*, B \in \mathcal{B}_{\mathbb{R}^n}$$

y definimos $T : E \to \mathbb{R}^n$ como $T(x) = (f_1(x), \cdots, f_n(x))$. De este modo, al componer T con v_i obtenemos dos vectores aleatorios en $\mathbb{R}^n$, a saber $T(v_i) = (f_1(v_i), \cdots, f_n(v_i))$. Por

Cramer-Wold sabemos que la distribución de estos está unívocamente determinada por la distribución de las combinaciones lineales de las coordenadas de dichos vectores. Es decir, dados $t_1, \ldots, t_n \in \mathbb{R}$ la distribución de $\sum_{i=1}^n t_i f_i(v_j)$ nos determina la distribución de $T(v_j)$ al variar $(t_1, \cdots, t_n) \in \mathbb{R}^n$. Como $\sum_{i=1}^n t_i f_i(v_j) = (t_1 f_1 + \cdots + t_n f_n)(v_j) = g(v_j)$ con $g = t_1 f_1 + \cdots t_n f_n \in E^*$, concluimos por hipótesis que $T(v_1)$ y $T(v_2)$ tienen la misma distribución en $\mathbb{R}^n$, es decir

$$P\left(T(v_1) \in B\right) = P\left(T(v_2) \in B\right)$$

Pero esto quiere decir que

$$P\left(v_1 \in A\right) = P\left((f_1(v_1), \cdots, f_n(v_1)) \in B\right) =$$

$$P\left((f_1(v_2), \cdots, f_n(v_2)) \in B\right) = P\left(v_2 \in A\right)$$

Luego $P_{v_1} = P_{v_2}$ en $\sigma(\mathcal{F})$. Como E es separable $\sigma(\mathcal{F}) = \mathcal{B}_E$. Es decir, v_1 y v_2 tienen la misma distribución. $\square$

Habíamos probado que $\mathcal{B}_E = \sigma(\{f^{-1}((-\infty, r]) : f \in E^*, r \in \mathbb{R}\})$. Luego las intersecciones finitas de los estos conjuntos forman un π-sistema que genera la σ-álgebra de Borel en E. Recordamos que un π-sistema es una colección de subconjuntos cerrada por intersecciones finitas. De la prueba de la proposición anterior podemos ver que si dos probabilidades coinciden en $\{f^{-1}((-\infty, r]) : f \in E^*, r \in \mathbb{R}\}$ -llamaremos semiespacios a estos conjuntos- entonces también coinciden en las intersecciones finitas. Luego la colección de semiespacios es una clase determinante para E cuando este es separable.

Proposición 1.10. *Sean E un espacio normado y $v_1, v_2 : \Omega \to E$ dos elementos aleatorios en E. Si v_1 y v_2 son independientes, entonces $f(v_1)$ y $g(v_2)$ son variables aleatorias independientes para todo $f, g \in E^*$. Más aún, si E es separable entonces el reciproco es cierto.*

Demostración. La primera parte de la proposición es inmediata y se deduce de la proposición 1.8. Supongamos que E es separable y que $f(v_1)$ y $g(v_2)$ son variables aleatorias independientes para todo $f, g \in E^*$. Entonces $E \times E$ es un espacio normado separable con $\mathcal{B}_{E\times E} = \mathcal{B}_E \times \mathcal{B}_E$, de donde (v_1, v_2) es un elemento aleatorio en $E \times E$. Además cada funcional $F \in (E \times E)^*$ es de la forma $F(x, y) = F(x, 0) + F(0, y) = f(x) + g(y)$ con $f, g \in E^*$. Luego para probar que v_1 y v_2 son independientes debemos probar que $P_{v_1,v_2} = P_{v_1} \times P_{v_2}$ -la medida producto en $E \times E$-. Como los semiespacios $\{(x, y) \in E \times E : F(x, y) \leq r\}$, $F \in (E \times E)^*$

y $r \in \mathbb{R}$ son una clase determinante, bastará probar que ambas medidas de probabilidad coinciden en estos semiespacios. Sea $A = \{(x, y) : f(x) + g(y) \leq r\}$ un semiespacio. Entonces

$$P_{v_1} \times P_{v_2}(A) = \int_E P_{v_2}(A_x) dP_{v_1}(x) = \int_E P\left(f(x) + g(v_2) \leq r\right) dP_{v_1}(x)$$

$$= \int_{\mathbb{R}} P\left(t + g(v_2) \leq r\right) dP_{f(v_1)}(t) = P_{f(v_1)} * P_{g(v_2)}((-\infty, r])$$

$$= P\left(f(v_1) + g(v_2) \leq r\right) = P_{v_1, v_2}(A)$$

Luego esta igualdad vale en todo $\mathcal{B}_{E\times E}$, es decir, v_1 y v_2 son independientes. □

1.4. Valor esperado de un elemento aleatorio.

En esta sección definiremos y estudiaremos el concepto de valor esperado o esperanza de un elemento aleatorio en un espacio normado separable. Nos concentraremos por ahora en definir tal concepto, demostrar unas propiedades básicas y probar un resultado de existencia para elementos aleatorios en espacios de Banach.

Definición 1.6. *Sean E un espacio normado separable y v un elemento aleatorio en E. Llamaremos valor esperado de v a un elemento $x \in E$ que cumpla*

$$f(x) = \mathrm{E}f(v) = \int_\Omega f(v(\omega)) dP(\omega)$$

para todo funcional $f \in E^*$. En caso de existir tal elemento lo denotaremos $x = \mathrm{E}v$.

Como E^* es una familia de funciones que separa puntos -es decir, si $x \neq 0 \in E$ entonces existe $f \in E^*$ tal que $f(x) \neq 0$- en caso de existir un elemento x que cumpla la definición anterior, debe ser único. Pues si x_1 es otro que la cumple, se tiene que

$$f(x_1) = \mathrm{E}f(v) = f(x) \quad \forall f \in E^*$$

de donde $x - x_1 = 0$.

De una forma similar definimos la varianza de un elemento aleatorio como el promedio de las distancias cuadráticas de v a $\mathrm{E}v$. Es decir,

Definición 1.7. *Sea v un elemento aleatorio en un espacio normado separable E y supongamos que existe su valor esperado* $\mathrm{E}v$. *Definimos la varianza de v como*

$$\sigma^2(v) = \int_\Omega \|v - \mathrm{E}v\|^2 dP$$

Llamaremos desvío estándar de v a $\sigma(v)$, la raíz cuadrada de la varianza.

Comencemos por ver algunas propiedades.

Proposición 1.11. *Sean E un espacio normado separable, v y w dos elementos aleatorios en E y $a \in E$ un elemento fijo. Entonces*

1. *Si existen* $\mathrm{E}v$ *y* $\mathrm{E}w$*, entonces* $\mathrm{E}v + w = \mathrm{E}v + \mathrm{E}w$.

2. *Si existe* $\mathrm{E}v$ *y* $r \in \mathbb{R}$*, entonces* $\mathrm{E}rv = r\mathrm{E}v$.

3. *Si* $v = a$ *con probabilidad 1, entonces* $\mathrm{E}v = a$.

4. *Si* $v = a$ *con probabilidad 1 y* α *es una variable aleatoria tal que* $\mathrm{E}\alpha$ *existe, entonces* $\mathrm{E}\alpha v = (\mathrm{E}\alpha)a$.

5. *Si* $T : E \to F$ *es un operador lineal acotado y* $\mathrm{E}v$ *existe, entonces* $\mathrm{E}T(v)$ *también existe y vale* $\mathrm{E}T(v) = T(\mathrm{E}v)$.

6. *Si* $\mathrm{E}v$ *existe entonces* $\|\mathrm{E}v\| \leq \mathrm{E}\,\|v\|$*, pudiendo esta última ser infinita.*

Demostración. 1. Sea $x = \mathrm{E}v + \mathrm{E}w$. Entonces

$$f(x) = f(\mathrm{E}v + \mathrm{E}w) = f(\mathrm{E}v) + f(\mathrm{E}w)$$

$$= \mathrm{E}f(v) + \mathrm{E}f(w) = \mathrm{E}f(v) + f(w) = \mathrm{E}f(v + w)$$

para todo $f \in E^*$.

2. Sea $x = r\mathrm{E}v$. Entonces

$$f(x) = f(r\mathrm{E}v) = rf(\mathrm{E}v) = r\mathrm{E}f(v) = \mathrm{E}rf(v) = \mathrm{E}f(rv)$$

para todo $f \in E^*$.

3. Tenemos que $f(a) = \mathrm{E}f(v)$ para todo $f \in E^*$, pues $f(v) = f(a)$ con probabilidad 1.

4. Tenemos que $f((\mathrm{E}\alpha)a) = (\mathrm{E}\alpha)f(a) = \mathrm{E}(\alpha f(a)) = \mathrm{E}(\alpha f(v)) = \mathrm{E}f(\alpha v)$ para todo $f \in E^*$.

5. Sea $y = T(\mathrm{E}v)$. Entonces si $f \in F^*$,

$$f(y) = f(T(\mathrm{E}v)) = f \circ T(\mathrm{E}v) = \mathrm{E}f \circ T(v) = \mathrm{E}f(T(v))$$

pues $f \circ T \in E^*$.

6. Sea $f \in E^*$ tal que $\|f\| = 1$ y $f(\mathrm{E}v) = \|\mathrm{E}v\|$. Entonces

$$\|\mathrm{E}v\| = f(\mathrm{E}v) = |f(\mathrm{E}v)| = |\mathrm{E}f(v)|$$

$$\leq \mathrm{E}|f(v)| \leq \mathrm{E}\,\|f\|\,\|v\| = \mathrm{E}\,\|v\|$$

$\square$

El siguiente resultado nos garantiza la existencia del valor esperado de un elemento aleatorio v en E, si $\mathrm{E}\,\|v\| < +\infty$ y E es completo. Es importante también pues en la prueba se calcula el valor esperado de elementos discretos en estas condiciones. Luego se obtiene la esperanza de v aproximando por elementos discretos de la misma forma que se hace en $\mathbb{R}$.

Proposición 1.12. *Sea E un espacio de Banach. Si v es un elemento aleatorio en E tal que $\mathrm{E}\,\|v\| < +\infty$, entonces $\mathrm{E}v$ existe.*

Demostración. Supongamos primero que v es discreto y que toma los valores $\{x_1, x_2, \ldots\}$. Entonces bajo estas hipótesis tenemos que

$$\mathrm{E}\,\|v\| = \sum_{n\geq 1} \|x_n\|\, P\left(E_n\right) < +\infty$$

donde $E_n = v^{-1}(\{x_n\})$. Luego $x_N = \sum_{n=1}^{N} x_n P\left(E_n\right)$ converge por la completitud de E al valor $x = \sum_{n\geq 1} x_n P\left(E_n\right)$. Pero entoces si $f \in E^*$

$$f(x) = f\left(\lim_N \sum_{n=1}^{N} x_n P\left(E_n\right)\right) = \lim_N \sum_{n=1}^{N} P\left(E_n\right) f(x_n)$$

y como $\sum_{n=1}^{N} P\left(E_n\right)|f(x_n)| \leq \|f\| \sum_{n=1}^{N} P\left(E_n\right)\|x_n\|$, tenemos que $\sum_{n=1}^{+\infty} P\left(E_n\right)|f(x_n)| < +\infty$, de donde

$$f(x) = \sum_{n=1}^{+\infty} P\left(E_n\right) f(x_n) = \mathrm{E}f(v)$$

Luego $\mathrm{E}v = \sum_{n=1}^{+\infty} P\left(E_n\right) x_n$.

En el caso general, sea $\{v_n\}$ una sucesión de elementos aleatorios discretos tales que $\|v_n - v\| < \frac{1}{n}$ para cada n. Por lo anterior sabemos que existe para cada n, $\mathrm{E}v_n \in E$. Probemos que $\{\mathrm{E}v_n\}$ es una sucesión de Cauchy. Sea $p \in \mathbb{N}$,

$$\|\mathrm{E}v_{n+p} - \mathrm{E}v_n\| = \|\mathrm{E}v_{n+p} - v_n\| \leq \mathrm{E}\,\|v_{n+p} - v_n\| \leq \frac{1}{n} + \frac{1}{n+p} \to 0 \ \ \forall p \in \mathbb{N}$$

Como E es completo existe $x \in E$ tal que $\lim_n \mathrm{E} v_n = x$. Luego

$$|f(x) - \mathrm{E} f(v)| = |\lim_n f(\mathrm{E} v_n) - \mathrm{E} f(v)| \leq \limsup_n |\mathrm{E} f(v_n) - f(v)|$$

$$\leq \limsup_n \mathrm{E} |f(v_n - v)| \leq \|f\| \lim_n \mathrm{E} \|v_n - v\| = 0$$

de donde $f(x) = \mathrm{E} f(v)$. Como esto vale para todo $f \in E^*$, concluímos que $x = \mathrm{E} v$. $\square$

La hipótesis de completitud es en efecto necesaria. Por ejemplo consideremos el espacio $E = \mathbb{R}^\infty$ formado por todas las sucesiones reales con sólo una cantidad finita de términos no nulos, con la métrica del supremo. Entonces E no es completo. Sea $v : \Omega \to \mathbb{R}^\infty$ tal que $P(v = e_n) = \frac{1}{2^n}$ donde $e_n = \{\delta_{in}\}_{i \geq 1} \in \mathbb{R}^\infty$. Entonces $\|v\| = 1$ con probabilidad 1, y por lo tanto $\mathrm{E} \|v\| = 1 < +\infty$. Pero si pensamos a v como un e.a. en c_0 -las sucesiones convergentes a cero- que es la clausura de $\mathbb{R}^\infty$, y por ende un espacio completo; entonces $\mathrm{E} v = \sum_{n=1}^{+\infty} \frac{1}{2^n} e_n = \{1/2, 1/4, 1/8, \ldots\} \in c_0$ pero no es un elemento de $\mathbb{R}^\infty$.

Consideremos el espacio $\mathbf{c}$ de las sucesiones reales convergentes. Entonces $\mathbf{c}$ es un espacio de Banach separable con la norma

$$\|x\| = \sup_n |x_n| \quad \text{para } x = \{x_n\} \in \mathbf{c}$$

Supongamos que $v = \{v_n\}$ es un e.a. en $\mathbf{c}$ tal que

$$(*) \quad \mathrm{E} \|v\| = \mathrm{E} \sup_n |v_n| < \infty$$

Afirmamos que $\mathrm{E} v = \{\mathrm{E} v_n\}$. Como existe $\lim_n v_n$ en todo Ω, por $(*)$ podemos aplicar convergencia dominada y concluir que también existe $\lim_n \mathrm{E} v_n = \mathrm{E} \lim_n v_n$, de donde $\{\mathrm{E} v_n\} \in \mathbf{c}$. Además, si $f \in \mathbf{c}^*$, entonces podemos identificar f con $(f_0, \{f_n\})$ donde $f_0 \in \mathbb{R}$ y $\{f_n\} \in l^1$. De este modo

$$f(\{\mathrm{E} v_n\}) = f_0 \lim_n \mathrm{E} v_n + \sum_{n \geq 1} f_n \mathrm{E} v_n = \mathrm{E} \left\{ f_0 \lim_n v_n + \sum_{n \geq 1} f_n v_n \right\} = \mathrm{E} f(v)$$

Consideremos ahora el espacio de Banach $\mathbf{C} = C[0,1]$ de las funciones continuas $x : [0,1] \to \mathbb{R}$ con la norma

$$\|x\| = \sup_{t \in [0,1]} |x(t)|$$

Sea v un e.a. en $\mathbf{C}$, tal que

$$\mathrm{E} \|v\| = \mathrm{E} \sup_t |v(t)| < \infty$$

Probaremos que $\mathrm{E}v = \{\mathrm{E}v_t : t \in [0,1]\}$. Primero observemos que la aplicación $t \mapsto \mathrm{E}v_t$ -que llamaremos x- es continua. En efecto tenemos que $v_{t+h} - v_t \to 0$ cuando $h \to 0$ y como $\mathrm{E}\,\|v\| < \infty$ nuevamente si aplicamos convergencia dominada vemos que

$$|\mathrm{E}v_{t+h} - \mathrm{E}v_t| \leq \mathrm{E}|v_{t+h} - v_t| \to 0$$

cuando $h \to 0$. Como $\mathbf{C}^*$ es el conjunto de las medidas signadas finitas en $[0,1]$, tenemos que dado $\mu \in \mathbf{C}^*$,

$$\mu(x) = \int_0^1 \mathrm{E}v_t d\mu(t) = \int_0^1 \int_\Omega v(t,\omega) dP(\omega) d\mu(t) = \mathrm{E}\left\{\int_0^1 v_t d\mu(t)\right\} = \mathrm{E}\mu(v)$$

por lo que $x = \mathrm{E}v$.

A continuación veamos cómo podemos calcular el valor esperado cuando el espacio en el que estamos trabajando tiene una base de Schauder. Sea E un espacio de Banach con base de Schauder $\{e_n\}$. Ponemos e_k^* a las funcionales coordenadas relativas a la base $\{e_n\}$, es decir, si

$$x = \sum_i x_i e_i$$

entonces $e_k^*(x) = x_k$. Como E es de Banach tenemos que -consecuencia del teorema de la aplicación abierta- $e_k^* \in E^*$. Luego si v es un elemento aleatorio en E, $e_k^*(v)$ es una variable aleatoria para todo k. Más aún, si consideramos P_n los operadores de sumas parciales

$$P_n(x) = \sum_{i=1}^n x_i e_i$$

estos también son continuos. Como $\{e_n\}$ es una base, para cada $x \in E$ fijo tenemos que

$$\|x - P_n(x)\| \to_n 0$$

es decir, $P_n(x) \to x$. Luego $P_n(v)$ converge puntualmente a v y podemos escribir

$$v = \lim_n P_n(v) = \sum_n e_n^*(v) e_n$$

Supongamos que existe $\mathrm{E}v$. Entonces debemos tener

$$e_k^*(\mathrm{E}v) = \mathrm{E}e_k^*(v)$$

de donde podemos escribir

$$\mathrm{E}v = \sum_n \mathrm{E}e_n^*(v) e_n$$

Luego si identificamos v con la sucesión $(v_1, v_2, \cdots)$ donde $v_n = e_n^*(v)$ podemos identificar a $\mathrm{E}v$ con

$$\mathrm{E}v = (\mathrm{E}v_1, \mathrm{E}v_2, \cdots)$$

Capítulo 2

Ley de los Grandes Números.

En este capítulo estudiaremos la ley fuerte de los grandes números para e.a.. En la primera parte nos concentraremos en e.a. a valores en un espacio de Hilbert separable. Probaremos las extensiones de las desigualdades de Rademacher-Mensov para el caso ortogonal y de Kolmogorov para el caso independiente. Luego probaremos en ambos casos las leyes fuertes que se deducen de éstas. La teoría en este caso transcurre del mismo modo que en el caso real y se utilizan los mismos métodos de prueba.

En la segunda parte veremos la ley fuerte de Mourier para el caso i.i.d. en espacios de Banach separables, analizaremos algunos ejemplos que invalidan las extensiones de la ley fuerte y débil para e.a. independientes con restricciones en sus varianzas.

2.1. Generalización a espacios de Hilbert separables.

2.1.1. Definiciones y aspectos básicos.

A lo largo de esta sección denotamos por H un espacio de Hilbert real, que suponemos separable, y denotamos por $\langle \cdot, \cdot \rangle$ su producto interno. De este modo tenemos definida la norma como $\|x\| = \sqrt{\langle x, x \rangle}$.

Si x e y son dos elementos aleatorios en H, entonces sabemos que $\|x - y\|$ es una variable aleatoria y más aún, tanto $x + y$ como $x - y$ son elementos aleatorios en H. Deducimos entonces que

$$\langle x, y \rangle = \frac{1}{4} \left(\|x + y\|^2 - \|x - y\|^2 \right)$$

es una variable aleatoria. Esto lo podríamos obtener diciendo que al ser H separable, $\mathcal{B}_{H\times H} = \mathcal{B}_H \times \mathcal{B}_H$; luego tenemos que (x, y) es un elemento aleatorio en $H \times H$ y como $\langle\ ,\ \rangle : H \times H \to \mathbb{R}$

es una función continua, $\langle x, y\rangle$ es una variable aleatoria.

Por la desigualdad de Markov que mencionamos en el capítulo anterior, al ser $d(x, y) = \|x - y\|$ tenemos que dado $\epsilon > 0$

$$P(\|x-y\| \geq \epsilon) \leq \frac{1}{\epsilon^r} \mathrm{E} \|x-y\|^r$$

si $\mathrm{E}\|x - y\|^r < +\infty$ para algún $r > 0$.

Definición 2.1. *Sean x e y dos elementos aleatorios en H con $\mathrm{E}\|x\|^2 < +\infty$ y $\mathrm{E}\|y\|^2 < +\infty$. Decimos que x e y son no correlacionados si*

$$\mathrm{E}\langle x, y\rangle = \langle \mathrm{E}x, \mathrm{E}y\rangle$$

Una familia arbitraria de elementos aleatorios es no correlacionada si dos cualquiera de ellos lo son.

Cuando

$$\mathrm{E}\langle x, y\rangle = 0$$

decimos que x e y son ortogonales. Así, una familia arbitraria de elementos aleatorios en H es ortogonal si dos cualquiera de ellos los son. Si ademas $\mathrm{E}\|x\|^2 = 1$ para todo elemento de la familia, decimos que es una familia ortonormal.

Si x e y son tal que $\mathrm{E}x = \mathrm{E}y = 0$, entonces x e y son no correlacionados si y sólo si x e y son ortogonales. Como

$$\mathrm{E}\langle x - \mathrm{E}x, y - \mathrm{E}y\rangle = \mathrm{E}\langle x, y\rangle - \mathrm{E}\langle x, \mathrm{E}y\rangle - \mathrm{E}\langle \mathrm{E}x, y\rangle + \langle \mathrm{E}x, \mathrm{E}y\rangle$$

y como $\langle \cdot, \mathrm{E}y\rangle$ y $\langle \mathrm{E}x, \cdot\rangle$ son elementos de H^*, de la definición de valor esperado obtenemos que $\mathrm{E}\langle x, \mathrm{E}y\rangle = \langle \mathrm{E}x, \mathrm{E}y\rangle$ y $\mathrm{E}\langle \mathrm{E}x, y\rangle = \langle \mathrm{E}x, \mathrm{E}y\rangle$. Entonces deducimos que

$$\mathrm{E}\langle x - \mathrm{E}x, y - \mathrm{E}y\rangle = \mathrm{E}\langle x, y\rangle - \langle \mathrm{E}x, \mathrm{E}y\rangle$$

Luego podemos afirmar que x e y son no correlacionados si y sólo si $x - \mathrm{E}x$ e $y - \mathrm{E}y$ son ortogonales.

Observemos que hemos exigido que los "momentos" segundos de x e y sean finitos. Esto no es formalmente necesario pero bajo estas condiciones tenemos que

$$\mathrm{E}|\langle x, y\rangle| \leq \mathrm{E}\|x\| \|y\| \leq \left(\mathrm{E}\|x\|^2\right)^{1/2} \left(\mathrm{E}\|y\|^2\right)^{1/2} < +\infty$$

por la desigualdad de Cauchy-Schwarz.

Los cálculos que hicimos recién nos permiten definir la covarianza de dos elementos aleatorios x e y en H como

$$\operatorname{cov}(x, y) = \mathrm{E}\langle x - \mathrm{E}x, y - \mathrm{E}y\rangle = \mathrm{E}\langle x, y\rangle - \langle \mathrm{E}x, \mathrm{E}y\rangle$$

y tenemos que dos elementos aleatorios son no correlacionados si y sólo si su covarianza es cero.

Habíamos definido la varianza de un elemento aleatorio x como

$$\sigma^2(x) = \int_\Omega \|x - \mathrm{E}x\|^2 \, d\mathrm{P} = \mathrm{E}\,\|x - \mathrm{E}x\|^2 = \operatorname{cov}(x, x)$$

y obtenemos como en el caso real,

$$\sigma^2(x) = \mathrm{E}\,\|x\|^2 - \|\mathrm{E}x\|^2$$

Además la desigualdad de Chebychev toma la siguiente forma:

Si x es un elemento aleatorio en H con varianza finita, entonces para todo $\epsilon > 0$

$$P(\|x - \mathrm{E}x\| \geq \epsilon) \leq \frac{1}{\epsilon^2}\sigma^2(x)$$

que se deduce de la desigualdad de Markov para $r = 2$.

Sean $x_1, x_2, \ldots, x_n$ elementos aleatorios en H. Entonces denotamos por $s_n = x_1 + \cdots + x_n$. Así, tenemos que

$$\sigma^2(s_n) = \mathrm{E}\langle s_n - \mathrm{E}s_n, s_n - \mathrm{E}s_n\rangle = \mathrm{E}\langle \sum_{i=1}^n x_i - \mathrm{E}x_i, \sum_{j=1}^n x_j - \mathrm{E}x_j\rangle$$

$$= \sum_{i=1}^n \sum_{j=1}^n \mathrm{E}\langle x_i - \mathrm{E}x_i, x_j - \mathrm{E}x_j\rangle = \sum_{i=1}^n \sigma^2(x_i) + 2 \sum_{1 \leq i < j \leq n} \operatorname{cov}(x_i, x_j)$$

Osea

$$\sigma^2(s_n) = \sum_{i=1}^n \sigma^2(x_i) + 2 \sum_{1 \leq i < j \leq n} \operatorname{cov}(x_i, x_j)$$

Así tenemos probado el siguiente lema:

Lema 2.1. *Si $x_1, \ldots, x_n$ son elementos aleatorias en H no correlacionados, entonces*

$$\sigma^2(s_n) = \sum_{i=1}^n \sigma^2(x_i)$$

La importancia de este lema la veremos más adelante y aunque es exactamente el mismo resultado que en el caso real, es justamente el resultado que permite extender la teoría del caso real a los espacios de Hilbert.

Antes de comenzar con la ley de los grandes números veremos una extensión de los lemas de Toeplitz y Kronecker para espacios de Banach, que usaremos más adelante.

Lema 2.2 (Lema de Toeplitz). *Sea* $\{u_{nk}\}_{n,k\geq 1}$ *una sucesión doble de números reales positivos tales que*

$$(1) \quad \lim_{n\to+\infty}\sum_{k=1}^{+\infty} u_{nk} = 1 \quad (2) \quad \lim_{n\to+\infty} u_{nk} = 0 \quad k = 1, 2, \ldots$$

Sea $s_1, s_2, \ldots$ *una sucesión en un espacio de Banach* $\mathcal{B}$ *que converge a* s*. Entonces*

$$t_n = \sum_{k=1}^{+\infty} u_{nk} s_k \to_n s$$

Demostración. Tenemos que

$$\|t_n - s\| = \left\|\sum_{k=1}^{+\infty} u_{nk} s_k - s\right\| = \left\|\sum_{k=1}^{+\infty} u_{nk}(s_k - s) + \left(\sum_{k=1}^{+\infty} u_{nk} - 1\right) s\right\|$$

$$\leq \left\|\sum_{k=1}^{k_0} u_{nk}(s_k - s)\right\| + \left\|\sum_{k=k_0+1}^{+\infty} u_{nk}(s_k - s)\right\| + \left|\sum_{k=1}^{+\infty} u_{nk} - 1\right| \|s\|$$

Fijemos k_0 tal que $\|s_k - s\| < \epsilon$ para todo $k \geq k_0$. Entonces

$$\left\|\sum_{k=k_0+1}^{+\infty} u_{nk}(s_k - s)\right\| \leq \sum_{k=k_0+1}^{+\infty} u_{nk} \|s_k - s\| < \epsilon \sum_{k=k_0+1}^{+\infty} u_{nk}$$

Si n_1 es tal que para todo $n \geq n_1$, $u_{nk} < \epsilon$ para $k = 1, 2, \ldots, k_0$, entonces tenemos que

$$\left\|\sum_{k=1}^{k_0} u_{nk}(s_k - s)\right\| \leq \sum_{k=1}^{k_0} u_{nk} \|s_k - s\| < k_0 \epsilon \sum_{k=1}^{k_0} \|s_k - s\|$$

y sea $n_0 \geq n_1$ tal que $\left|\sum_{k=1}^{+\infty} u_{nk} - 1\right| < \epsilon$ para todo $n \geq n_0$. Entonces

$$\left|\sum_{k=1}^{+\infty} u_{nk} - 1\right| \|s\| < \epsilon \|s\|$$

y como

$$\sum_{k=k_0+1}^{+\infty} u_{nk} \leq \left|\sum_{k=1}^{+\infty} u_{nk} - 1\right| + \sum_{k=1}^{k_0} u_{nk} + 1 < 1 + \epsilon + \epsilon k_0$$

tenemos que para todo $n \geq n_0$

$$\|t_n - s\| < \epsilon \left(\sum_{k=1}^{k_0} \|s_k - s\| + 1 + \epsilon(k_0+1) + \|s\| \right) = \epsilon M(k_0)$$

donde $M(k_0)$ es una constante fijado el k_0. Luego $t_n \to_n s$. $\square$

Lema 2.3 (Lema de Kronecker). *Sean $\{x_n\}$ una sucesión en un espacio de Banach $\mathcal{B}$ y $\{\lambda_n\}$ una sucesión creciente a infinito de números reales positivos. Si $\sum_{n=1}^{+\infty} \frac{x_n}{\lambda_n}$ converge en $\mathcal{B}$ entonces,*

$$\frac{x_1 + \cdots + x_n}{\lambda_n} \to_n 0$$

Demostración. Sea $s_n = \sum_{k=1}^{n} \frac{x_k}{\lambda_k}$ y $s = \sum_{k=1}^{+\infty} \frac{x_k}{\lambda_k}$, es decir $\|s_n - s\| \to 0$. Entonces tenemos que $s_n - s_{n-1} = \frac{x_n}{\lambda_n}$ para $n \geq 2$ y si ponemos $s_0 = 0$ y $\lambda_0 = 0$ podemos hacerlo para $n \geq 1$,

$$x_n = \lambda_n(s_n - s_{n-1})$$

Así tenemos que

$$\frac{1}{\lambda_n}\sum_{k=1}^{n} x_k = \frac{1}{\lambda_n}\sum_{k=1}^{n} \lambda_k(s_k - s_{k-1})$$

$$= \frac{1}{\lambda_n}\left(-s_0\lambda_1 + \sum_{k=1}^{n-1}(\lambda_k - \lambda_{k+1})s_k + \lambda_n s_n \right)$$

$$= s_n - \sum_{k=0}^{n-1} \frac{\lambda_{k+1} - \lambda_k}{\lambda_n} s_k$$

Entonces si ponemos $u_{nk} = \frac{\lambda_{k+1} - \lambda_k}{\lambda_n} > 0$ para $k = 0, 1, \ldots, n-1$ y $u_{nk} = 0$ para $k > n-1$, tenemos que

$$\sum_{k=0}^{+\infty} u_{nk} = \frac{1}{\lambda_n}\sum_{k=0}^{n-1} \lambda_{k+1} - \lambda_k = 1$$

$$u_{nk} = \frac{\lambda_{k+1} - \lambda_k}{\lambda_n} \to_n 0$$

pues $\lambda_n \to +\infty$. Luego por el lema de Toeplitz, tenemos que

$$t_n = \sum_k u_{nk} s_k = \sum_{k=0}^{n-1} \frac{\lambda_{k+1} - \lambda_k}{\lambda_n} s_k \to_n s$$

y entonces

$$\frac{1}{\lambda_n}\sum_{k=1}^{n} x_k = s_n - \sum_{k=0}^{n-1} \frac{\lambda_{k+1} - \lambda_k}{\lambda_n} s_k \to_n s - s = 0$$

$\square$

Corolario 2.1. *Sea $\{x_n\}$, una sucesión de elementos aleatorios en un espacio de Banach separable $\mathcal{B}$. Si*

$$\sum_{n=1}^{+\infty} \frac{x_n}{n}$$

es convergente en $\mathcal{B}$ con probabilidad 1, entonces

$$\frac{x_1 + \cdots + x_n}{n} \to_n 0 \quad c.s.$$

2.1.2. Sucesiones de elementos aleatorios no correlacionados.

Como vamos a trabajar con la esperanza del máximo -en norma- de las sumas parciales de e.a. ortogonales, nuestra próxima proposición es una útil desigualdad similar en su importancia a la desigualdad de Kolmogorov para elementos independientes, que nos servirá para probar la ley fuerte para e.a. no correlacionados.

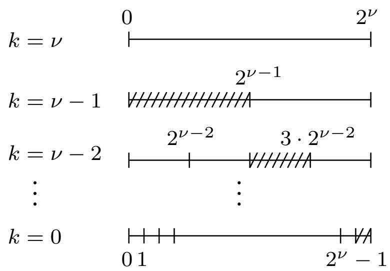
Proposición 2.1 (Desigualdad de Rademacher-Mensov). *Sean $x_1, x_2, \ldots, x_n$ elementos aleatorios ortonormales en H y $c_1, c_2, \ldots, c_n$ números reales. Entonces*[1]

$$\mathrm{E} \max_{1 \le k \le n} \left\| \sum_{j=1}^{k} c_j x_j \right\|^2 \le (\log^2 4n) \sum_{j=1}^{n} c_j^2$$

Demostración. Consideremos el caso $n = 2^\nu$ para un entero ν. Pongamos $s_j = c_1 x_1 + \cdots + c_j x_j$ y $z_{\alpha\beta} = c_{\alpha+1} x_{\alpha+1} + \cdots + c_\beta x_\beta$ para $\alpha = u2^k$, $\beta = (u+1)2^k$ con $k = 0, 1, \ldots, \nu$ y $u = 0, 1, \ldots, 2^{\nu-k} - 1$. Entonces cada

$$(1) \quad s_j = \sum_i z_{\alpha_i \beta_i}$$

para cierta cantidad de las $z_{\alpha\beta}$ de forma que podemos elegir $\beta_1 - \alpha_1 > \beta_2 - \alpha_2 > \cdots$.



[1] El logaritmo es en base 2.

Por ejemplo, s_n es la suma de las $z_{\alpha\beta}$ señaladas en el dibujo. De este modo la cantidad de sumandos en (1) es menor o igual que ν y tenemos por Cauchy-Schwarz,

$$\|s_j\|^2 = \left(\left\|\sum_i z_{\alpha_i\beta_i}\right\|\right)^2 \leq \left(\sum_i \|z_{\alpha_i\beta_i}\|\right)^2 \leq \nu\sum_i \|z_{\alpha_i\beta_i}\|^2 \leq \nu\sum_{\alpha,\beta}\|z_{\alpha\beta}\|^2$$

donde la última suma es tomada en todos los valores posibles de α y β. Como esto vale para todo j tenemos que

$$(2)\quad \max_{1\leq k\leq n}\left\|\sum_{j=1}^{k} c_j x_j\right\|^2 = \max_{1\leq k\leq n}\|s_k\|^2 \leq \nu\sum_{\alpha,\beta}\|z_{\alpha\beta}\|^2$$

Además

$$\mathrm{E}\nu\sum_{\alpha,\beta}\|z_{\alpha\beta}\|^2 = \nu\sum_{\alpha,\beta}\mathrm{E}\,\|z_{\alpha\beta}\|^2 = \nu\sum_{\alpha,\beta}c_{\alpha+1}^2+\cdots+c_\beta^2$$

donde la última igualdad se desprende de la ortonormalidad de los elementos. Como la última suma es tomada en todos los valores de α y β, separando en los valores de k desde 0 hasta ν y en cada una de ellas sumando de $u2^k$ hasta $(u+1)2^k$ para los valores $u = 0,\ldots,2^{\nu-k}-1$; obtenemos

$$\sum_{\alpha,\beta}c_{\alpha+1}^2+\cdots+c_\beta^2 = \sum_{k=0}^{\nu}\sum_{u=0}^{2^{\nu-k}-1}c_{u2^k+1}^2+\cdots+c_{(1+u)2^k}^2 = \sum_{k=0}^{\nu}\sum_{j=1}^{2^\nu}c_j^2 = (\nu+1)\sum_{j=1}^{2^\nu}c_j^2$$

de donde

$$\mathrm{E}\nu\sum_{\alpha,\beta}\|z_{\alpha\beta}\|^2 = \nu(\nu+1)\sum_{j=1}^{n}c_j^2 \leq (\nu+1)^2\sum_{j=1}^{n}c_j^2 = \log^2 2n\sum_{j=1}^{n}c_j^2$$

y tomando esperanzas en (2) obtenemos

$$\mathrm{E}\max_{1\leq k\leq n}\left\|\sum_{j=1}^{k}c_j x_j\right\|^2 \leq \log^2 2n\sum_{j=1}^{n}c_j^2$$

En general, sea ν tal que $2^\nu \leq n < 2^{\nu+1}$. Entonces ponemos $c_{n+1} = \cdots = c_{2^{\nu+1}} = 0$ y obtenemos que $s_k = s_n$ para $k \geq n$ y

$$\max_{1\leq k\leq 2^{\nu+1}}\|s_k\|^2 = \max_{1\leq k\leq n}\|s_k\|^2$$

de donde

$$\mathrm{E}\max_{1\leq k\leq n}\|s_k\|^2 = \mathrm{E}\max_{1\leq k\leq 2^{\nu+1}}\|s_k\|^2 \leq \log^2 2^{\nu+2}\sum_{j=1}^{2^{\nu+1}}c_j^2 \leq \log^2 4n\sum_{j=1}^{n}c_j^2$$

□

Teorema 2.1 (Rademacher-Mensov). *Sea $\{x_n\}$ una sucesión de elementos aleatorios en H ortonormales y $\{c_n\}$ una sucesión de números reales tales que*

$$\sum_{k=1}^{+\infty} c_k^2 \log^2 k < +\infty$$

entonces

$$\sum_{k=1}^{+\infty} c_k x_k$$

converge con probabilidad 1.

Demostración. Sea $R_n = \sum_{k=n}^{+\infty} c_k x_k$. Entonces por la ortonormalidad de los x_k tenemos que para todo $m \geq n$, $\mathrm{E}\,\|\sum_{k=n}^{m} c_k x_k\|^2 = \sum_{k=n}^{m} c_k^2 \leq \sum_{k=1}^{\infty} c_k^2 < \infty$ de donde tomando límite,

$$\mathrm{E}\,\|R_n\|^2 = \sum_{k=n}^{+\infty} c_k^2$$

Como

$$A = \sum_{k=1}^{+\infty} c_k^2 \log^2 k \geq \sum_{k=n}^{+\infty} c_k^2 \log^2 k \geq \sum_{k=n}^{+\infty} c_k^2 \log^2 n$$

tenemos que

$$\sum_{k=n}^{+\infty} c_k^2 \leq \frac{A}{\log^2 n}$$

por lo que

$$\mathrm{E}\,\|R_{2^n}\|^2 \leq \frac{A}{n^2}$$

lo cual implica que $R_{2^n} \to_n 0$ c.s.. Sea

$$D_n = \max_{2^n < k \leq 2^{n+1}} \|R_k - R_{2^n}\|$$

Entonces tenemos que

$$\mathrm{E}D_n^2 = \mathrm{E} \max_{2^n < k \leq 2^{n+1}} \left\| \sum_{j=2^n}^{k-1} c_j x_j \right\|^2 \leq \log^2(4(2^{n+1} - 2^n)) \sum_{j=2^n}^{2^{n+1}-1} c_j^2$$

$$= (n+2)^2 \sum_{j=2^n}^{2^{n+1}-1} c_j^2$$

pero este último es equivalente a

$$n^2 \sum_{j=2^n}^{2^{n+1}-1} c_j^2 = \log^2 2^n \sum_{j=2^n}^{2^{n+1}-1} c_j^2 \leq \sum_{j=2^n}^{2^{n+1}-1} c_j^2 \log^2 j$$

Luego, $\sum_{n=1}^{+\infty} \mathrm{E}D_n^2 < +\infty$ lo cual implica que $D_n \to 0$ c.s.. Sea $2^n \leq k < 2^{n+1}$, entonces

$$\|R_k\| \leq \|R_k - R_{2^n}\| + \|R_{2^n}\| \leq D_n + \|R_{2^n}\| \to 0 \quad c.s.$$

es decir, $R_k \to_k 0$ c.s. □

Teorema 2.2. *Sea $\{x_n\}$ una sucesión de elementos aleatorios ortogonales en H tales que $\mathrm{E}x_i = 0$ para todo i. Si*

$$\sum_{k=1}^{+\infty} \frac{\mathrm{E}\,\|x_k\|^2}{k^2} \log^2 k < +\infty$$

entonces

$$\frac{x_1 + \cdots + x_n}{n} \to 0 \quad c.s.$$

Demostración. Consideremos $y_n = \frac{x_n}{(\mathrm{E}\|x_n\|^2)^{1/2}}$ si $\mathrm{E}\,\|x_n\|^2 \neq 0$ e $y_n = 0$ si no[2]. Entonces

$$\mathrm{E}\langle y_i, y_j\rangle = \frac{1}{(\mathrm{E}\,\|x_i\|^2)^{1/2}} \frac{1}{(\mathrm{E}\,\|x_j\|^2)^{1/2}} \mathrm{E}\langle x_i, x_j\rangle = \delta_{ij}$$

es decir, y_n es una sucesión ortonormal de elementos aleatorios en H. Consideremos entonces

$$c_k = \frac{(\mathrm{E}\,\|x_k\|^2)^{1/2}}{k}$$

y tenemos por hipótesis que

$$\sum_{k=1}^{+\infty} c_k^2 \log^2 k = \sum_{k=1}^{+\infty} \frac{\mathrm{E}\,\|x_k\|^2}{k^2} \log^2 k < +\infty$$

Luego por el teorema de Rademacher-Mensov tenemos que

$$\sum_{k=1}^{+\infty} c_k y_k = \sum_{k=1}^{+\infty} \frac{(\mathrm{E}\,\|x_k\|^2)^{1/2}}{k} \frac{x_k}{(\mathrm{E}\,\|x_k\|^2)^{1/2}} = \sum_{k=1}^{+\infty} \frac{x_k}{k}$$

converge con probabilidad 1 y el lema de Kronecker nos permite afirmar la tesis. □

Teorema 2.3. *Sea $\{x_n\}$ una sucesión de elementos aleatorios no correlacionados en H tal que*

$$\sum_{k=1}^{+\infty} \frac{\sigma^2(x_k)}{k^2} \log^2 k < +\infty$$

entonces

$$\frac{s_n - \mathrm{E}s_n}{n} \to 0 \quad c.s.$$

Demostración. Basta observar que $y_n = x_n - \mathrm{E}x_n$ es una sucesión ortogonal, tal que $\mathrm{E}y_n = 0$ y $\mathrm{E}\,\|y_n\|^2 = \sigma^2(x_n)$. □

[2] En este caso se tiene $x_n = 0$ c.s..

2.1.3. Sucesiones de elementos aleatorios independientes.

Antes de demostrar los teoremas de esta sección probaremos el siguiente lema,

Lema 2.4. *Sean x e y dos elementos aleatorios en H tales que* $\mathrm{E}\left\|x\right\|^2 < +\infty$ *y* $\mathrm{E}\left\|y\right\|^2 < +\infty$. *Si x e y son independientes, entonces son no correlacionados.*

Demostración. Como x e y son no correlacionados si y sólo si $y-\mathrm{E}y$ y $x-\mathrm{E}x$ son ortogonales, podemos suponer que $\mathrm{E}x = \mathrm{E}y = 0$ y probar que x e y son ortogonales. Como H es un espacio de Hilbert separable, existe una base ortonormal completa $\{e_n\}$ en la cual podemos escribir el producto interno

$$\langle x, y\rangle = \sum_{n=1}^{+\infty}\langle x, e_n\rangle\langle y, e_n\rangle$$

Como $\|x\|\,\|y\| \in L^1(\Omega)$ y

$$\left|\sum_{n=1}^{m}\langle x, e_n\rangle\langle y, e_n\rangle\right| \leq \sum_{n=1}^{m}|\langle x, e_n\rangle|\,|\langle y, e_n\rangle| \leq$$

$$\left(\sum_{n=1}^{m}|\langle x, e_n\rangle|^2\right)^{1/2}\left(\sum_{n=1}^{m}|\langle y, e_n\rangle|^2\right)^{1/2} \leq \|x\|\,\|y\|$$

podemos aplicar el teorema de convergencia dominada para escribir

$$\mathrm{E}\langle x, y\rangle = \lim_{m\to+\infty}\sum_{n=1}^{m}\mathrm{E}\langle x, e_n\rangle\langle y, e_n\rangle$$

Pero $\langle\cdot, e_n\rangle \in H^*$ de donde $\langle x, e_n\rangle$ y $\langle y, e_n\rangle$ son independientes y

$$\mathrm{E}\langle x, e_n\rangle\langle y, e_n\rangle = \mathrm{E}\langle x, e_n\rangle\mathrm{E}\langle y, e_n\rangle = \langle\mathrm{E}x, e_n\rangle\langle\mathrm{E}y, e_n\rangle = 0$$

es decir, $\mathrm{E}\langle x, y\rangle = 0$ como queríamos demostrar. □

Teorema 2.4 (Desigualdad de Kolmogorov). *Sean $x_1, x_2, \ldots, x_n$ elementos aleatorios independientes en H. Supongamos además que* $\mathrm{E}x_k = 0$ *y* $\sigma^2(x_k) = \mathrm{E}\left\|x_k\right\|^2 < +\infty$, *para* $k = 1, \ldots, n$. *Entonces para todo $\epsilon > 0$*

$$P\left(\max_{1\leq k\leq n}\left\|\sum_{i=1}^{k}x_i\right\| > \epsilon\right) \leq \frac{1}{\epsilon^2}\sum_{i=1}^{n}\sigma^2(x_i)$$

Demostración. Fijemos un $\epsilon > 0$. Pongamos $\Lambda = \left\{ \text{máx}_{1\leq k\leq n} \left\| \sum_{i=1}^{k} x_i \right\| > \epsilon \right\}$. Para $k = 1, 2, \ldots, n$ sea

$$\Lambda_k = \{\|s_1\| \leq \epsilon, \ldots, \|s_{k-1}\| \leq \epsilon, \|s_k\| > \epsilon\}$$

donde denotamos $s_j = x_1 + \cdots + x_j$. Entonces es claro que $\bigcup_{k=1}^{n} \Lambda_k = \Lambda$, más aún una unión disjunta. Además

$$\int_{\Lambda} \|s_n\|^2 \, d\mathrm{P} = \sum_{k=1}^{n} \int_{\Lambda_k} \|s_n\|^2 \, d\mathrm{P} = \sum_{k=1}^{n} \int_{\Lambda_k} \|s_k + s_n - s_k\|^2 \, d\mathrm{P}$$

$$= \sum_{k=1}^{n} \int_{\Lambda_k} \|s_k\|^2 \, d\mathrm{P} + \sum_{k=1}^{n} \int_{\Lambda_k} \|s_n - s_k\|^2 \, d\mathrm{P} + 2 \sum_{k=1}^{n} \int_{\Lambda_k} \langle s_k, s_n - s_k \rangle d\mathrm{P}$$

Pero como

$$\int_{\Lambda_k} \langle s_k, s_n - s_k \rangle d\mathrm{P} = \int_{\Omega} 1_{\Lambda_k} \langle s_k, s_n - s_k \rangle d\mathrm{P} = \int_{\Omega} \langle 1_{\Lambda_k} s_k, s_n - s_k \rangle d\mathrm{P}$$

$$= \mathrm{E}\langle 1_{\Lambda_k} s_k, s_n - s_k \rangle = \langle \mathrm{E}1_{\Lambda_k} s_k, \mathrm{E}s_n - s_k \rangle = 0$$

Tenemos que

$$\sum_{i=1}^{n} \sigma^2(x_i) = \sigma^2(s_n) = \mathrm{E}\,\|s_n\|^2 = \int_{\Omega} \|s_n\|^2 \, d\mathrm{P} \geq \int_{\Lambda} \|s_n\|^2 \, d\mathrm{P}$$

$$= \sum_{k=1}^{n} \int_{\Lambda_k} \|s_k\|^2 \, d\mathrm{P} + \sum_{k=1}^{n} \int_{\Lambda_k} \|s_n - s_k\|^2 \, d\mathrm{P} \geq \sum_{k=1}^{n} \int_{\Lambda_k} \|s_k\|^2 \, d\mathrm{P}$$

$$\geq \epsilon^2 \sum_{k=1}^{n} P(\Lambda_k) = \epsilon^2 P(\Lambda)$$

de donde resulta la desigualdad. □

Teorema 2.5. *Sea $\{x_n\}$ una sucesión de elementos aleatorios independientes en H. Si*

$$\sum_{n=1}^{+\infty} \frac{\sigma^2(x_n)}{n^2} < +\infty$$

entonces

$$\frac{1}{n} \sum_{i=1}^{n} x_i - \mathrm{E}x_i \to_n 0 \quad c.s.$$

Demostración. Primero observamos que alcanza con probar el caso en que $\mathrm{E}x_n = 0$ para todo n. Denotamos con $s_n = x_1 + \cdots x_n$, y definimos

$$M_n = \max_{2^n < k \leq 2^{n+1}} \frac{\|s_k\|}{k}$$

Probaremos que $M_n \to 0$ c.s.. Dado $\epsilon > 0$, tenemos que

$$P\left(M_n > \epsilon\right) = P\left(\max_{2^n < k \leq 2^{n+1}} \frac{\|s_k\|}{k} > \epsilon\right) \leq P\left(\max_{2^n < k \leq 2^{n+1}} \|s_k\| > \epsilon 2^n\right)$$

$$\leq P\left(\max_{k \leq 2^{n+1}} \|s_k\| > \epsilon 2^n\right) \leq \frac{1}{\epsilon^2 4^n} \sum_{k=1}^{2^{n+1}} \sigma^2\left(x_k\right)$$

donde la última desigualdad se deduce de la desigualdad de Kolmogorov. Entonces

$$\sum_{n=1}^{\infty} P\left(M_n > \epsilon\right) \leq \sum_{n=1}^{\infty} \frac{1}{\epsilon^2 4^n} \sum_{k=1}^{2^{n+1}} \sigma^2\left(x_k\right) = \epsilon^{-2} \sum_{k=1}^{\infty} \sigma^2\left(x_k\right) \sum_{n:2^{n+1} \geq k} \frac{1}{4^n}$$

Para cada entero $k \geq 1$, existe un único entero $j(k)$ tal que $2^{j(k)} \leq k \leq 2^{j(k)+1}$. Luego

$$\sum_{n:2^{n+1} \geq k} \frac{1}{4^n} = \sum_{n=j(k)}^{\infty} \frac{1}{4^n} = \frac{4}{3 \cdot 4^{j(k)}} \leq \frac{16}{3k^2}$$

De este modo

$$\sum_{n=1}^{\infty} P\left(M_n > \epsilon\right) \leq \frac{16}{3\epsilon^2} \sum_{k=1}^{\infty} \frac{\sigma^2\left(x_k\right)}{k^2} < \infty$$

Entonces por Borel-Cantelli deducimos que $M_n \to 0$ c.s.. □

2.2. Generalización a espacios de Banach separables.

Cuando trabajamos en un espacio de Banach vemos que la situación es drásticamente diferente. En los espacios de Hilbert no tuvimos ningún inconveniente en extender prácticamente todos los resultados relativos a la ley fuerte de los grandes números que conocemos del caso real, y más aún los métodos que utilizamos para probarlos fueron exactamente los mismos. Se puede decir, que esto se debe al importante y simple resultado 2.1, la varianza de la suma de e.a. no correlacionados es la suma de las varianzas.

Para espacios de Banach es muy poco lo que se puede decir en un contexto general. Por ejemplo, no de forma inmediata pero sí relativamente simple se puede extender la ley fuerte de Kolmogorov para variables independientes con igual distribución, ley que probó Mourier en 1953. Las principales desventajas que aparecen en demostrar leyes para e.a. no correlacionados radican en la carencia de un producto natural en el espacio, teniendo que recurrirse a definiciones en el sentido débil. Pero sin ir tan lejos, la ley fuerte para e.a. independientes no es cierta -en general- bajo la única hipotesis de varianzas uniformemente acotadas. Tampoco es verdadera la ley fuerte de Kolmogorov -caso independientes sólo- que probamos en la sección anterior, ni siquiera para espacios uniformemente convexos y reflexivos. Antes de probar el teorema de Mourier veamos algunos ejemplos de todo esto.

Ejemplos

Consideremos el espacio l^1 de las sucesiones reales absolutamente sumables. Sean $\delta^n \in l^1$ dado por $\delta^n = \{\delta_{in}\}_i$ y $\{A_n\}$ una sucesión de variables aleatorias reales tales que con probabilidad 1/2 toman los valores 1 y -1. Entonces definimos

$$v_n = A_n \delta^n = \{0, \ldots, 0, A_n, 0, \ldots\}$$

Como podemos identificar $(l^1)^*$ con l^∞, dado $f = \{f_i\} \in (l^1)^*$ tenemos que

$$f(v_n) = \sum_i f_i A_n \delta_{in} = f_n A_n$$

por lo que $\{f(v_n)\}$ es una sucesión de elementos aleatorios independientes para todo $f \in (l^1)^*$, de donde $\{v_n\}$ es una sucesión de e.a. independientes. Además tenemos que $\mathrm{E}v_n = 0$ y como

$$\|v_n\| = |A_n| = 1$$

deducimos que $\sigma^2(v_n) = 1$ para todo n. Sin embargo, también tenemos que

$$\|v_1 + \cdots + v_n\| = |A_1| + \cdots + |A_n| = n$$

de donde $\sigma^2(v_1 + \cdots + v_n) = n^2$ y vemos que no se cumple el lema 2.1. Más aún

$$\left\| \frac{1}{n} \sum_{k=1}^n v_k \right\| = \frac{1}{n} \sum_{k=1}^n |A_k| = 1 \quad \forall n$$

por lo que $n^{-1} \sum_1^n v_k$ no converge a cero.

Consideremos ahora el espacio l^p con $1 < p < 2$. Sea $1 - p^{-1} < q < 1/2$. Denotamos igual que antes δ^n y sea $\{A_n\}$ una sucesión de variables aleatorias reales independientes que toman los valores $\pm n^q$ con probabilidad $1/2$. Tomamos como antes $v_n = A_n \delta^n$ que forman una sucesión de e.a. independientes en l^p. Así, $\mathrm{E} v_n = 0$ para cada n y

$$\|v_n\| = \left(\sum_i |A_n \delta_{in}|^p \right)^{1/p} = |A_n| = n^q$$

de donde $\sigma^2(v_n) = n^{2q}$. Luego

$$\sum_n \frac{\sigma^2(v_n)}{n^2} = \sum_n \frac{n^{2q}}{n^2} = \sum_n n^{2q-2} < \infty$$

Pero

$$\left\| \frac{1}{n} \sum_{k=1}^n v_n \right\|^p = \frac{1}{n^p} \sum_{k=1}^n |A_k|^p = \sum_{k=1}^n \frac{k^{pq}}{n^p} > \sum_{k=[n/2]}^n \frac{k^{pq}}{n^p} > \sum_{k=[n/2]}^n \frac{(n/2)^{pq}}{n^p} > \frac{n^{pq-p+1}}{2^{pq+1}} \to \infty$$

pues $q > 1 - p^{-1}$.

Teorema 2.6 (de Mourier). *Sea $\{v_n\}$ una sucesión de elementos aleatorios independientes e identicamente distribuidos en un espacio de Banach separable X. Si $\mathrm{E}\|v_1\| < +\infty$ entonces*

$$\frac{1}{n} \sum_{i=1}^n v_i \to_n \mathrm{E} v_1 \quad c.s.$$

Demostración. Como $\mathrm{E}\|v_1\| < +\infty$ y X es un espacio de Banach separable, tenemos que existe $\mathrm{E} v_1$. Además como v_n y v_1 son identicamente distribuidos tenemos que $\mathrm{E} v_n = \mathrm{E} v_1$ para todo n.

Supongamos primero que los v_n son discretos tomando una cantidad numerable de valores $\{a_n\}$. Para cada entero $k \geq 1$ definimos

$$v_n^k = \begin{cases} v_n & \text{si } v_n \in \{a_1, \ldots, a_k\} \\ 0 & \text{si no.} \end{cases}$$

Entonces tenemos que $\{v_n^k\}$ son una sucesión de elementos aleatorios independientes e identicamente distribuidos tales que

$$\mathrm{E} \left\| v_1^k \right\| = \sum_{i=1}^k \|a_i\| P(v_1 = a_i) \leq \mathrm{E}\|v_1\| < +\infty$$

y con la particularidad siguiente: toman valores en un espacio de Banach de dimensión finita, a saber $V = \langle \{a_1, \ldots, a_k\} \rangle$. Luego, para cada k existe $A_k^1 \subseteq \Omega$ de probabilidad 1 tal que

$$\frac{1}{n} \sum_{i=1}^n v_i^k \to_n \mathrm{E} v_1^k$$

Sea también para cada k

$$r_n^k = v_n - v_n^k$$

Entonces $\{\left\|r_n^k\right\|\}$ es una sucesión de variables aleatorias independientes e identicamente distribuidas con

$$\mathrm{E}\left\|r_1^k\right\| = \sum_{i=k+1}^{+\infty} \|a_i\| P(v_1 = a_i) \leq \mathrm{E}\|v_1\| < +\infty$$

Luego para cada k existe $A_k^2 \subseteq \Omega$ de probabilidad 1 en donde

$$\frac{1}{n}\sum_{i=1}^{n}\left\|r_i^k\right\| \to_n \mathrm{E}\left\|r_1^k\right\|$$

Como $r_1^k \to_k 0$ puntualmente en Ω y $\mathrm{E}\left\|r_1^k\right\| \leq \mathrm{E}\|v_1\|$ para todo k tenemos que

$$\mathrm{E}\left\|r_1^k\right\| \to_k 0$$

Entonces, sean $\epsilon > 0$ y k tal que

$$\mathrm{E}\left\|r_1^k\right\| < \epsilon$$

Consideremos el conjunto $A = \bigcap_k (A_k^1 \cap A_k^2)$ también de probabilidad 1. Luego en A vale

$$\left\|\frac{1}{n}\sum_{i=1}^{n} v_i - \mathrm{E}v_1\right\| \leq \frac{1}{n}\sum_{i=1}^{n}\left\|r_i^k\right\| + \left\|\frac{1}{n}\sum_{i=1}^{n} v_i^k - \mathrm{E}v_1^k\right\| + \mathrm{E}\left\|r_1^k\right\|$$

Como el primer término converge al tercero que es menor que ϵ, podemos elegir n suficientemente grande para que su suma sea menor que 3ϵ. Para el término del medio, como k es fijo por lo que observamos antes podemos tomar n grande para que sea menor que ϵ. Luego

$$\left\|\frac{1}{n}\sum_{i=1}^{n} v_i - \mathrm{E}v_1\right\| < 4\epsilon$$

para $n \geq n_0$.

En la situación general, como X es separable, para todo $\lambda > 0$ existe $T_\lambda : X \to X$ Borel medible que toma una cantidad numerable de valores y tal que $\|T_\lambda x - x\| \leq \lambda$ para todo $x \in X$. Entonces tenemos que para cada λ

$$\left\|\frac{1}{n}\sum_{i=1}^{n} v_i - \mathrm{E}v_1\right\| \leq \frac{1}{n}\sum_{i=1}^{n}\|T_\lambda v_i - v_i\| + \left\|\frac{1}{n}\sum_{i=1}^{n} T_\lambda v_i - \mathrm{E}T_\lambda v_1\right\| + \mathrm{E}\|T_\lambda v_1 - v_1\|$$

$$\leq 2\lambda + \left\|\frac{1}{n}\sum_{i=1}^{n} T_\lambda v_i - \mathrm{E}T_\lambda v_1\right\|$$

Por lo que hemos probado el segundo término converge a cero cuando $n \to +\infty$ en un conjunto A_λ de probabilidad 1 para cada $\lambda > 0$. Sea $\lambda_m = \frac{1}{m}$ y sea $A = \bigcap_m A_{1/m}$. Dado un $\epsilon > 0$ arbitrario fijemos m tal que $\frac{1}{m} < \epsilon$, entonces en el conjunto A vale

$$\left\| \frac{1}{n} \sum_{i=1}^{n} v_i - \mathrm{E} v_1 \right\| \leq 2\epsilon + \left\| \frac{1}{n} \sum_{i=1}^{n} T_{1/m} v_i - \mathrm{E} T_{1/m} v_1 \right\| < 3\epsilon$$

si $n \geq n_0$. □

Cuando se retira la hipótesis de identica distribución vimos que los teoremas fallan para espacios de Banach en general. En el primer ejemplo la situación es un poco diferente que la del segundo, ya que se puede probar que si el espacio cumple una condición de convexidad, es válida la ley fuerte para e.a. independientes con varianzas uniformemente acotadas. La condición de la que hablamos es la condición de Beck. Un espacio de Banach se dice del tipo (B) si existen un entero positivo k y un $\epsilon > 0$ tales que para todo $x_1, \ldots, x_k$ con norma menor o igual que 1 se tiene

$$\|\pm x_1 \pm \cdots \pm x_k\| > k(1 - \epsilon)$$

para alguna elección de los signo $+$ y $-$. Dentro de esta clase se encuentran por ejemplo, los espacios uniformemente convexos.

Capítulo 3

Teorema Central del Límite.

Estudiaremos ahora la convergencia en distribución de e.a., principalmente a valores en un espacio de Hilbert. El capítulo tiene como objetivo principal el teorema central del límite para e.a. indepenidentes e identicamente distribuídos en un espacio de Hilbert separable. Para ello, utilizamos el método de las funciones características, probamos la existencia de e.a. gaussianos utilizando el teorema de Minlos-Sazonov. Estudiamos también las propiedades de compacidad relativa de una familia de medidas de probabilidad en espacios métricos mediante el teorema de Prohorov y damos un criterio para ésta en espacios de Hilbert.

3.1. Convergencia débil en espacios métricos.

Sea (M, d) un espacio métrico y $\mathcal{B}$ la σ-álgebra de Borel de M. Diremos que una sucesión $\{P_n\}$ de medidas de probabilidad en M converge débilmente a P -también medida de probabilidad en M- si

$$\int_M f(x)dP_n(x) \to_n \int_M f(x)dP(x)$$

para toda función $f : M \to \mathbb{R}$ continua y acotada. En símbolos escribiremos $P_n \Rightarrow P$. Hay muchas otras formas de definir este concepto, por supuesto todas equivalentes que no probaremos aquí. Utilizaremos las siguientes:

- $P_n \Rightarrow P$.
- $P(F) \geq \limsup_n P_n(F)$ para todo $F \subseteq M$ cerrado.
- $P(G) \leq \liminf_n P_n(G)$ para todo $G \subseteq M$ abierto.
- $P_n(A) \to_n P(A)$ para todo $A \in \mathcal{B}$ tal que $P(\partial A) = 0$.

Los conjuntos que cumplan $P(\partial A) = 0$ los llamaremos conjuntos de P-continuidad o simplemente de continuidad.

El concepto de tensión es uno de los más importantes que manejaremos. Dada una familia arbitraria de medidas de probabilidad Π en M, diremos que es tensa si para cada $\epsilon > 0$ existe un compacto $K \subset M$ tal que

$$P(K) > 1 - \epsilon$$

para toda $P \in \Pi$. La importancia de este concepto radica en el siguiente teorema de Prohorov.

Teorema 3.1. *Sea Π una familia de medidas de probabilidad en M. Si Π es tensa, entonces Π es relativamente compacta.*

Demostración. Sea $\{P_n\}$ una sucesión en Π. Debemos probar que existe una subsucesión que converge débilmente a una medida de probabilidad P no necesariamente perteneciente a Π. Elegimos $K_1 \subseteq M$ compacto de forma tal que $P_n(K_1) > 0$ para todo n. A continuación, elegimos $\hat{K}_2 \subseteq M$ tal que $P_n(\hat{K}_2) > 1 - \frac{1}{2}$ para todo n. Sea $K_2 = K_1 \cup \hat{K}_2$ que también es compacto. De este modo, obtenemos una sucesión creciente $\{K_u\}$ de subconjuntos compactos tales que

$$P_n(K_u) > 1 - \frac{1}{u}, \quad \forall n, u$$

Tomamos $K = \bigcup_u K_u$ -puede no ser compacto-. Entonces K es un subconjunto separable de M pues cada K_u lo es al ser compacto. Así, existe una familia $\mathcal{A}$ numerable de abiertos -de hecho bolas abiertas- con la siguiente propiedad: para todo abierto G de M y todo punto $x \in K \cap G$ existe un abierto $A \in \mathcal{A}$, $x \in A \subset \overline{A} \subset G$. Sea $\mathcal{H}$ formada por el conjunto vacío y la familia numerable de las uniones finitas de conjuntos de la forma $\overline{A} \cap K_u$ con $A \in \mathcal{A}$ y $u \geq 1$.

Como para cada $H \in \mathcal{H}$ la sucesión $\{P_n(H)\}$ está acotada y $\mathcal{H}$ es numerable, aplicando un razonamiento diagonal podemos extraer una subsucesión $\{P_{n_i}\}$ tal que exista el límite

$$\alpha(H) = \lim_i P_{n_i}(H)$$

para todo $H \in \mathcal{H}$. La idea es a partir de α extender primero a todos los abiertos, para luego construir una medida exterior que restringida a los borelianos funcione bien.

Veamos primero qué propiedades cumple α. Si $H_1 \subseteq H_2$, entonces como

$$\alpha(H_1) = \lim P_{n_i}(H_1) \leq \lim P_{n_i}(H_2) = \alpha(H_2)$$

α es monótona; este mismo argumento nos muestra que α es subaditiva en uniones finitas, aditiva en uniones finitas disjuntas y $\alpha(\emptyset) = 0$.

Para cada abierto G de M definimos

$$\beta(G) = \sup_{H \subseteq G} \alpha(H)$$

Así, tenemos que $\beta(\emptyset) = \alpha(\emptyset) = 0$. Si $G_1 \subseteq G_2$, como todo H que esté incluído en G_1 lo estará también en G_2, tenemos que β es monótona. A continuación definimos para cada subconjunto S de M

$$\gamma(S) = \inf_{S \subseteq G} \beta(G)$$

donde el ínfimo lo tomamos en todos los abiertos que contengan a S. Veamos entonces que γ es una medida exterior.

Tenemos por un lado que $\beta(G) \geq \gamma(G)$ para todo G abierto, pues siempre éste es un abierto que se contiene a si mismo. Pero por la monotonía de β debe ser $\gamma(G) = \beta(G)$; es decir γ y β coinciden en los abiertos. Así, en particular $\gamma(\emptyset) = \beta(\emptyset) = 0$. Además, si $S_1 \subseteq S_2$ todo abierto que contenga a S_2 también contendrá a S_1, y en consecuencia γ es también monótona. Resta probar que γ es subaditiva en uniones numerables.

Sean G_1 y G_2 abiertos. Sea $H \in \mathcal{H}$ tal que $H \subseteq G_1 \cup G_2$. Definimos los conjuntos cerrados

$$F_1 = \{x \in H : d(x, G_1^c) \geq d(x, G_2^c)\}$$

$$F_2 = \{x \in H : d(x, G_2^c) \geq d(x, G_1^c)\}$$

Es claro que F_1 y F_2 son cerrados. Sea $x \in F_1$, si x estuviera en G_1^c claramente tendríamos $d(x, G_1^c) = 0$ y por lo tanto $d(x, G_2^c) = 0$. Al ser éste cerrado tendríamos que $x \in G_2^c$ pero esto implicaría que $x \in (G_1 \cup G_2)^c \subseteq H^c$ lo cual es absurdo. Luego $F_1 \subseteq G_1$ y del mismo modo probamos que $F_2 \subseteq G_2$. Observemos además que $H \subseteq F_1 \cup F_2$.

El paso siguiente será encontrar subconjuntos H_1 y H_2 de $\mathcal{H}$ que estén entre medio de F_i y G_i, es decir, que $F_i \subseteq H_i \subseteq G_i$. Procedamos con F_1. Como $H \in \mathcal{H}$,

$$H = \bigcup_{j=1}^{l} \overline{A_j} \cap K_{u_j}$$

para ciertos $A_j \in \mathcal{A}$ y enteros u_j. Sea $u = \max\{u_j\}$, entonces $F_1 \subseteq K_u$ que es compacto por lo que F_1 también lo es. Para cada punto $x \in F_1$, como $F_1 \subseteq G_1$ podemos encontrar $A_x \in \mathcal{A}$ tal que $x \in A_x \subseteq \overline{A_x} \subseteq G_1$. Luego existen $x_1, \ldots, x_r$ en F_1 tales que

$$F_1 \subseteq \bigcup_{j=1}^{r} \overline{A_{x_j}}$$

Sea entonces $H_1 = \bigcup_{j=1}^{r} \overline{A_{x_j}} \cap K_u$. Del mismo modo procedemos con F_2.

Luego tenemos que

$$\alpha(H) \leq \alpha(H_1 \cup H_2) \leq \alpha(H_1) + \alpha(H_2) \leq \beta(G_1) + \beta(G_2)$$

y tomando supremo en H tenemos $\beta(G_1 \cup G_2) \leq \beta(G_1) + \beta(G_2)$. Podemos concluir entonces que β es finitamente subaditiva.

Sea $\{G_k\}$ una sucesión de abiertos arbitraria. Tomamos $H \subseteq \bigcup_k G_k$ con $H \in \mathcal{H}$. Como H es compacto deben existir $G_{k_1}, \ldots, G_{k_N}$ tales que $H \subseteq \bigcup_{j=1}^{N} G_{k_j}$. Entonces

$$\alpha(H) \leq \beta\left(\bigcup_{j=1}^{N} G_{k_j}\right) \leq \sum_{j=1}^{N} \beta(G_{k_j}) \leq \sum_k \beta(G_k)$$

y tomando supremo en H obtenemos que

$$\beta\left(\bigcup_k G_k\right) \leq \sum_k \beta(G_k)$$

Podemos probar ahora que γ es σ-subaditiva. Sea $\{S_k\}$ una sucesión de subconjuntos de M. Dado $\epsilon > 0$ arbitrario, para cada k elegimos $G_k \supseteq S_k$ abierto tal que $\beta(G_k) < \gamma(S_k) + \frac{\epsilon}{2^k}$. Luego como $G = \bigcup_k G_k$ es un abierto que contiene a $\bigcup_k S_k$ tenemos que

$$\gamma\left(\bigcup_k S_k\right) \leq \beta\left(\bigcup_k G_k\right) \leq \sum_k \beta(G_k) \leq \sum_k \gamma(S_k) + \epsilon$$

Como ϵ es arbitrario concluímos que γ es σ-subaditiva. Es decir, efectivamente γ es una medida exterior.

Probemos ahora que todo cerrado es γ-medible. Sean $F \subseteq M$ un cerrado, $G \subseteq M$ un abierto y $\epsilon > 0$ dado. Sea $H_0 \in \mathcal{H}$ tal que $H_0 \subseteq G \cap F^c$ y $\beta(G \cap F^c) < \alpha(H_0) + \epsilon$. Elegimos ahora $H_1 \in \mathcal{H}$ tal que $H_1 \subset G \cap H_0^c$ y $\beta(G \cap H_0^c) < \alpha(H_1) + \epsilon$. Entonces tenemos que

$$\gamma(G \cap F) + \gamma(G \cap F^c) \leq \beta(G \cap H_0^c) + \beta(G \cap F^c) \leq \alpha(H_1) + \alpha(H_0) + 2\epsilon$$

$$= \alpha(H_0 \cup H_1) + 2\epsilon \leq \beta(G) + 2\epsilon$$

Luego

$$\beta(G) \geq \gamma(G \cap F) + \gamma(G \cap F^c) \text{ para todo } G \text{ abierto.}$$

Si $L \subseteq M$ es un conjunto arbitrario, entonces como para todo $L \subseteq G$ abierto se tiene que

$$\gamma(L \cap F) + \gamma(L \cap F^c) \leq \gamma(G \cap F) + \gamma(G \cap F^c) \leq \beta(G)$$

tomando ínfimo en G obtenemos la desigualdad que buscabamos. Como los conjuntos γ-medibles son una σ-álgebra que contiene a los cerrados, contiene a los borelianos. Además al ser γ una medida cuando la restringimos a los conjuntos γ-medibles, podemos obtener una medida definida en los borelianos simplemente con la restricción de γ a estos. Además como $K_u \in \mathcal{H}$ para todo u[1] se tiene

$$1 \geq \gamma(M) = \beta(M) = \sup_{H \subseteq M} \alpha(H) \geq \sup_u \alpha(K_u) \geq \sup_u 1 - \frac{1}{u} = 1$$

de donde $P = \gamma|_{\mathcal{B}}$ es una medida de probabilidad.

Además, para todo abierto G se tiene que

$$P(G) = \beta(G) = \sup_{H \subseteq G} \alpha(H)$$

Como $P_{n_i}(H) \leq P_{n_i}(G)$ para todo n_i tenemos que

$$\alpha(H) = \lim_i P_{n_i}(H) \leq \liminf_i P_{n_i}(G) \quad \forall H \subseteq G$$

y en conclusión se obtiene que

$$P(G) \leq \liminf_i P_{n_i}(G)$$

para todo abierto G de donde $P_{n_i} \Rightarrow P$. □

Teorema 3.2. *Sea M un espacio métrico completo y separable. Si Π es una familia de medidas de probabilidad en M relativamente compacta, entonces Π es tensa.*

Demostración. Fijemos $\epsilon > 0$. Probemos primero lo siguiente: si $\{G_n\}$ es una sucesión creciente a M de abiertos, entonces existe n tal que $P(G_n) > 1 - \epsilon$ para toda $P \in \Pi$. Si no fuera así, tendríamos para cada n una $P_n \in \Pi$ para la cual $P_n(G_n) \leq 1 - \epsilon$. Como Π es relativamente compacta debe existir Q medida de probabilidad y una subsucesión $\{P_{n_i}\}$ tales que $P_{n_i} \Rightarrow Q$. Pero como $Q(M) = \lim Q(G_n)$ y para cada $m \leq n$ tenemos $P_n G_m \leq 1 - \epsilon$, vale para todo n fijo

$$Q(G_n) \leq \liminf_i P_{n_i}(G_n) \leq 1 - \epsilon$$

de donde $Q(M) \leq 1 - \epsilon$ lo cual es absurdo. Consideremos los conjuntos $A_{ik} = B(x_i, 1/k)$ donde $\{x_k\}$ es un conjunto denso numerable en M. Entonces para cada k tenemos que $\bigcup_i A_{ik} = M$, por lo que debe existir $n = n(k)$ tal que

$$P(\bigcup_{i=1}^{n(k)} A_{ik}) > 1 - \frac{\epsilon}{2^k}$$

[1] Como son compactos están incluidos en una unión finita de elementos de $\mathcal{A}$.

para toda $P \in \Pi$. Luego sea $A = \bigcap_k \bigcup_{i=1}^{n(k)} A_{ik}$. Entonces A es un conjunto totalmente acotado y como M es completo, $K = \overline{A}$ es compacto. Pero además,

$$P(K^c) \leq P(A^c) \leq \sum_k P\left((\bigcup_{i=1}^{n(k)} A_{ik})^c \right) < \epsilon$$

es decir, K es el compacto que buscabamos. $\square$

A continuación probaremos un teorema también de Prohorov que nos da un criterio para determinar la tensión de una familia de probabilidades en un espacio de Hilbert separable. Sea entonces H de Hilbert separable y sea $\{e_i\}$ una base ortonormal completa de H. Entonces para cada $x \in H$ y para cada $N \geq 1$ definimos

$$r_N^2(x) = \sum_{i=N}^{\infty} \langle x, e_i \rangle^2 = \sum_{i=N}^{\infty} x_i^2$$

donde denotamos $x_i = \langle x, e_i \rangle$.

Teorema 3.3. *Sea Π una familia arbitraria de medidas de probabilidad en H. Si se cumple*

$$\lim_{N\to\infty} \sup_{P\in\Pi} \int_H r_N^2(x) dP(x) = 0$$

entonces Π es tensa.

Demostración. Denotemos $\psi_N = \sup_{P\in\Pi} \int_H r_N^2(x) dP(x)$. Es claro que $\{\psi_N\}$ es decreciente, y la hipótesis del teorema es que $\psi_N \searrow 0$. Fijemos $\epsilon > 0$. Entonces podemos encontrar -pasando si es necesario por una subsucesión- una sucesión $0 < \alpha_N \to \infty$ tal que $\sum_N \alpha_N \psi_N \leq \epsilon$. Definimos entonces los conjuntos

$$K_N = \{x \in H : r_N^2(x) \leq \alpha_N^{-1}\}$$

y consideremos $K = \bigcap_N K_N$. Como $K \subseteq K_1 = \{x : \|x\|^2 \leq \alpha_1^{-1}\}$ tenemos que K es débilmente -secuencialmente- compacto. Probemos que K es efectivamente compacto en norma. Consideremos $\{x_m\}$ una sucesión en K. Entonces existe una subsucesión $\{x_{m'}\}$ y un $x \in K$ tales que $x_{m'} \to x$ débilmente. Pero como

$$\|x_{m'} - x\|^2 = \sum_i \langle x_{m'} - x, e_i \rangle^2 = \sum_{i=1}^{N-1} \langle x_{m'} - x, e_i \rangle^2 + \sum_{i=N}^{\infty} \langle x_{m'} - x, e_i \rangle^2$$

Para cada i tenemos que

$$\langle x_{m'} - x, e_i \rangle^2 = \langle x_{m'}, e_i \rangle^2 + \langle x, e_i \rangle^2 - 2\langle x_{m'}, e_i \rangle \langle x, e_i \rangle \leq$$

$$\langle x_{m'}, e_i\rangle^2 + \langle x, e_i\rangle^2 + 2|\langle x_{m'}, e_i\rangle||\langle x, e_i\rangle|$$

De donde sumando tenemos que

$$\sum_{i=N}^{\infty}\langle x_{m'} - x, e_i\rangle^2 \leq r_N^2(x_{m'}) + r_N^2(x) + 2\sum_{i=N}^{\infty}|\langle x_{m'}, e_i\rangle||\langle x, e_i\rangle|$$

$$\leq r_N^2(x_{m'}) + r_N^2(x) + 2\left(\sum_{i=N}^{\infty}\langle x_{m'}, e_i\rangle^2\right)^{1/2}\left(\sum_{i=N}^{\infty}\langle x, e_i\rangle^2\right)^{1/2}$$

$$= (r_N(x_{m'}) + r_N(x))^2 \leq \frac{4}{\alpha_N}$$

Luego tenemos que

$$\|x_{m'} - x\|^2 \leq \sum_{i=1}^{N-1}\langle x_{m'} - x, e_i\rangle^2 + \frac{4}{\alpha_N}$$

Entonces eligiendo primero N para que $\frac{4}{\alpha_N} < \frac{\epsilon}{2}$ y luego haciendo m' suficientemente grande para que la suma finita sea también menor que $\epsilon/2$ tenemos que

$$\|x_{m'} - x\|^2 < \epsilon$$

si m' es suficientemente grande.

Por último vemos por la desigualdad de Chebychev que para toda $P \in \Pi$

$$P(K^c) \leq \sum_N P(K_N^c) \leq \sum_N \alpha_N \int_H r_N^2(x)dP(x) \leq \sum_N \alpha_N \psi_N < \epsilon$$

$\square$

3.2. Funciones características.

Nuevamente nos concentramos en un espacio de Banach X separable y sea X^* su espacio dual. Consideremos una medida de probabilidad en $\mathcal{B}$, la σ-álgebra de Borel de X. Definiremos la transformada de Fourier de P que denotaremos por $\hat{P} : X^* \to \mathbb{C}$ como

$$\hat{P}(l) = \int_X e^{il(x)}dP(x) \quad \forall l \in X^*$$

De este modo tenemos que $\hat{P}(0) = 1$ y que $|\hat{P}(l)| \leq 1$ para todo $l \in X^*$. Además como

$$|\hat{P}(l+h) - \hat{P}(l)| \leq \int_M |e^{ih(x)} - 1|dP(x) \leq \int_M \text{mín}\{\|h\|\,\|x\|\,, 2\}dP(x)$$

Y el integrando en la derecha está acotado por 2 y converge a cero cuando $h \to 0$ para cada x, aplicando el teorema de la convergencia dominada vemos que $\hat{P}$ es uniformemente continua.

También vemos fácilmente que $\hat{P}(-l) = \overline{\hat{P}(l)}$ y que si P_1 y P_2 son dos medidas de probabilidad en X, denotando $P_1 * P_2$ la convolución tenemos que

$$P_1 \hat{*} P_2(l) = \hat{P}_1(l)\hat{P}_2(l) \quad \forall l \in X^*$$

Por último mencionamos que al igual que en el caso finito dimensional tenemos que $\hat{P}$ es definida positiva.

Proposición 3.1 (Unicidad). *Sean P_1 y P_2 dos medidas de probabilidad en X. Si $\hat{P}_1 = \hat{P}_2$ entonces $P_1 = P_2$.*

Demostración. Sea $l \in X^*$ fijo. Consideremos $l : X \to \mathbb{R}$ como una aplicación medible y definamos $\mu_i = P_i(l^{-1})$, es decir para todo boreliano A de $\mathbb{R}$ tenemos que $\mu_i(A) = P_i(l^{-1}(A))$. Entonces como $tl \in X^*$ para todo $t \in \mathbb{R}$ tenemos por la fórmula de cambio de variable que

$$\hat{P}_i(tl) = \int_X e^{itl(x)} dP_i(x) = \int_{\mathbb{R}} e^{itu} d\mu_i(u) = \hat{\mu}_i(t)$$

Pero entonces $\hat{\mu}_1 = \hat{\mu}_2$ y por el teorema de unicidad para $\mathbb{R}$ debe ser $\mu_1 = \mu_2$. Pero como los semi-espacios son una clase determinante debemos tener $P_1 = P_2$. □

Proposición 3.2 (Continuidad). *Supongamos que $\{P_n\}$ es una sucesión de medidas de probabilidad en X tal que $P_n \Rightarrow P$, entonces $\hat{P}_n(l) \to \hat{P}(l)$ para todo $l \in X^*$. Reciprocamente, si $\{P_n\}$ es una sucesión tensa de medidas de probabilidad tal que $\hat{P}_n(l) \to \theta(l)$ para todo $l \in X^*$, para cierta función $\theta : X^* \to \mathbb{C}$, entonces existe P medida de probabilidad en X tal que $P_n \Rightarrow P$ y $\hat{P} = \theta$.*

Demostración. Si $P_n \Rightarrow P$ entonces como $\hat{P}_n(l)$ es la integral de una función continua y acotada la conclusión la deducimos de la definición.

Reciprocamente, como $\{P_n\}$ es tensa existe una subsucesión $\{P_{n_i}\}$ y una medida de probabilidad P, tal que $P_{n_i} \Rightarrow P$. Pero entonces por lo que recién probamos tenemos que $\hat{P}_{n_i}(l) \to \hat{P}(l)$ para todo $l \in X^*$. Como también se tiene $\hat{P}_{n_i}(l) \to \theta(l)$ debe ser $\theta = \hat{P}$. Pero este argumento muestra que lo mismo ocurre para cualquier subsucesión. Es decir, toda subsucesión que converja débilmente converge en efecto a una misma medida límite, a saber P, debido a la unicidad de la transformada de Fourier. Luego necesariamente $P_n \Rightarrow P$. □

Si x es un elemento aleatorio en X definimos la función característica de x, $\varphi_x : X^* \to \mathbb{C}$ como la transformada de Fourier de su distribución. Es decir

$$\varphi_x(l) = \int_X e^{il(y)} dP_x(y) = Ee^{il(x)}$$

Por último, observemos que si H es un espacio de Hilbert separable, la correspondencia que nos brinda el teorema de Riesz entre H y H^* dada por $h \mapsto \langle \cdot, h \rangle$ establece un isomorfismo lineal isométrico. Entonces podemos pensar la transformada de Fourier de una medida de probabilidad en H como una aplicación $\hat{P} : H \to \mathbb{C}$ dada por

$$\hat{P}(h) = \int_H e^{i\langle x,h \rangle} dP(x)$$

y ésta conserva las propiedades que recién probamos.

3.3. Teorema de Minlos-Sazonov.

En esta sección demostraremos el teorema de Minlos-Sazonov que es una generalización para espacios de Hilbert separables del teorema de Bochner. Es decir, se plantea el siguiente problema: dada una función $\varphi : H \to \mathbb{C}$ continua con $\varphi(0) = 1$, cuándo es φ una función característica? En el caso finito dimensional es conocido el teorema de Bochner que afirma que una condición necesaria y suficiente para que φ sea una función característa es que sea definida positiva. El resultado de Minlos-Sazonov lo utilizaremos para poder definir un elemento gaussiano en H. Para probarlo necesitamos algunas herramientas.

Sea H un espacio de Hilbert real y separable. Denotamos $\mathcal{B}$ la σ-álgebra de Borel en H. El lema que sigue básicamente lo hemos probado en la proposición 1.6 de la sección 1.3, pero para hacer esta sección más autocontenida lo enunciaremos y probaremos de nuevo en una forma ligeramente distinta.

Dado un subespacio finito dimensional $\mathcal{U}$ de H denotamos $\mathcal{B}_\mathcal{U}$ la σ-álgebra de Borel en $\mathcal{U}$. Sea $\pi_\mathcal{U}$ la proyección ortogonal sobre $\mathcal{U}$, entonces diremos que un subconjunto $E \subseteq H$ es un cilindro de base $\mathcal{U}$ si existe $B \in \mathcal{B}_\mathcal{U}$ tal que $E = \pi_\mathcal{U}^{-1} B$. Llamamos $\mathcal{B}^\mathcal{U}$ la σ-álgebra de todos los cilindros de base $\mathcal{U}$. Por último sea $\mathcal{B}^0 = \bigcup_\mathcal{U} \mathcal{B}^\mathcal{U}$. Como a un cilidro siempre se le puede ampliar la base, es fácil ver que $\mathcal{B}^0$ es un álgebra. La continuidad de las proyecciones nos garantiza que todo cilindro es boreliano y de hecho vale el siguiente lema.

Lema 3.1. *$\mathcal{B}$ coincide con $\sigma(\mathcal{B}^0)$.*

Demostración. Basta ver que toda bola cerrada pertenece a $\sigma(\mathcal{B}^0)$. Sea

$$S = \{x \in H : \|x - x_0\| \leq \rho\}$$

Consideremos una base ortonormal $\{e_i\}$ y sea $\mathcal{U}_n = \langle e_1, \ldots, e_n \rangle$. Entonces definimos

$$S_n = \{x : \|\pi_n(x - x_0)\| \leq \rho\}$$

es decir, $S_n = \pi_n^{-1} B_n$ donde B_n es la bola cerrada de centro $\pi_n(x_0)$ y radio ρ en $\mathcal{U}_n$. Luego cada $S_n \in \mathcal{B}^0$. Es claro que como $\|\pi_n(x - x_0)\| \leq \|x - x_0\|$ se tiene que $S \subseteq S_n$ para todo n. Veamos que $\bigcap_n S_n \subseteq S$.

De hecho, si $x \notin S$ entonces existe $\epsilon > 0$ tal que

$$\|x - x_0\| = \rho + \epsilon$$

Como $\pi_n(x - x_0) \to x - x_0$, para n suficientemente grande se tiene necesariamente que $\|\pi_n(x - x_0)\| > \rho$. $\square$

Dada una medida de probabilidad P en $(H, \mathcal{B})$ podemos definir una familia de probabilidades $\{P_{\mathcal{U}}\}$ en $\{(\mathcal{U}, \mathcal{B}_{\mathcal{U}})\}$ de la siguiente manera. Como $\pi_{\mathcal{U}}$ es una función continua de H en $\mathcal{U}$ podemos escribir para cada $B \in \mathcal{B}_{\mathcal{U}}$

$$P_{\mathcal{U}}(B) = P(\pi_{\mathcal{U}}^{-1} B)$$

A las $P_{\mathcal{U}}$ las llamaremos proyecciones de P.

Definición 3.1. *Sea $\{Q_{\mathcal{U}}\}$ una familia de medidas de probabilidad en $\{(\mathcal{U}, \mathcal{B}_{\mathcal{U}})\}$. Decimos que es una familia compatible si para todo $\mathcal{U}_1 \subset \mathcal{U}_2$ y todo $E \in \mathcal{B}_{\mathcal{U}_1}$ se tiene que*

$$Q_{\mathcal{U}_1}(E) = Q_{\mathcal{U}_2}\left(\pi_{\mathcal{U}_1}^{-1} E \cap \mathcal{U}_2\right)$$

Es claro por como definimos las proyecciones de una medida de probabilidad en H que éstas forman una familia compatible. Nuestra siguente proposición dará una condición necesaria y suficiente para que dada una familia compatible de medidas finito-dimensionales, exista una medida P cuyas proyecciones coincidan con la familia original.

Proposición 3.3. *Sea $\{P_{\mathcal{U}}\}$ una familia compatible de medidas de probabilidad finito-dimensionales en H. Entonces existe una probabilidad P definida en H tal que las proyecciones de P son las $P_{\mathcal{U}}$ si y sólo si para todo $\epsilon > 0$ existe N_0 tal que si $N \geq N_0$ entonces,*

$$\sup P_{\mathcal{U}}\left(\{\|x\| \geq N\} \cap \mathcal{U}\right) < \epsilon$$

donde el supremo es tomado en todos los subespacios $\mathcal{U}$ de dimensión finita.

Demostración. De existir dicha P entonces como $P(\{\|x\| \geq N\}) \rightarrow_N 0$, dado $\epsilon > 0$ existe N_0 tal que si $N \geq N_0$ entonces

$$P(\{\|x\| \geq N\}) < \epsilon$$

Luego dado $\mathcal{U}$ de dimension finita tenemos que

$$P_{\mathcal{U}}(\{\|x\| \geq N\} \cap \mathcal{U}) = P(\pi_{\mathcal{U}}^{-1}\{\{\|x\| \geq N\} \cap \mathcal{U}\}) \leq P(\{\|x\| \geq N) < \epsilon$$

Probemos el reciproco. Como la familia $\{P_{\mathcal{U}}\}$ es compatible podemos definir sin ambigüedad la función de conjuntos en $\mathcal{B}^0$ siguiente; si $E \in \mathcal{B}^0$ entonces existe $\mathcal{U}$ de dimensión finita y $B \in \mathcal{B}_{\mathcal{U}}$ tal que $E = \pi_{\mathcal{U}}^{-1}B$. Entonces ponemos

$$P(E) = P_{\mathcal{U}}(B)$$

De este modo vemos que P es positiva, finitamente aditiva -pues a todo cilindro se le puede agrandar la base- y $P(H) = 1$. Como $\mathcal{B}^0$ es un álgebra que genera los borelianos bastará probar que P es σ-aditiva en $\mathcal{B}^0$ o lo que es equivalente, la continuidad por arriba en el vacío.

Sea entonces $\{E_n\}$ una sucesión decreciente al vacío de conjuntos en $\mathcal{B}^0$. Probaremos que $P(E_n) \downarrow 0$. Para cada n sea $\mathcal{U}_n$ un subespacio de dimensión finita con $E_n \in \mathcal{B}_{\mathcal{U}_n}$ y podemos suponer sin pérdida de generalidad que $\mathcal{U}_n$ es creciente -nuevamente esto es pues a todo cilindro se le puede agrandar la base-. Para cada n sabemos que existe $B_n \in \mathcal{B}_{\mathcal{U}_n}$ tal que $E_n = \pi_n^{-1}B_n$, donde π_n denota la proyección ortogonal sobre $\mathcal{U}_n$.

Supongamos primero que cada B_n es un subconjunto compacto por lo tanto cerrado y acotado de $\mathcal{U}_n$. En este caso, como las proyecciones son débilmente continuas tenemos que cada E_n es débilmente cerrado. Sea

$$S_N = \{x : \|x\| \leq N\}$$

Entonces sabemos que S_N es un conjunto débilmente compacto y por lo tanto también lo es para todo n, $S_N \cap E_n$. Además como $S_N \cap E_n \downarrow \emptyset$ por la compacidad debe existir un $n_0 = n_0(N)$ tal que $S_N \cap E_n = \emptyset$ para todo $n \geq n_0$. Dado $\epsilon > 0$, sea $N = N(\epsilon)$ tal que

$$\sup P_{\mathcal{U}}\left(\{\|x\| \geq N\} \cap \mathcal{U}\right) < \epsilon$$

Luego por lo anterior, existe $n_0 = n_0(N) = n_0(\epsilon)$ tal que $E_n \subseteq \{x : \|x\| > N\}$ si $n \geq n_0$. Como $B_n \subseteq E_n$ tenemos que

$$P(E_n) = P_{\mathcal{U}_n}(B_n) \leq P_{\mathcal{U}_n}(\{\|x\| > N\} \cap \mathcal{U}_n) < \epsilon$$

Veamos el caso general. Sea $\epsilon > 0$ y para cada n sea $C_n \subseteq B_n$ compacto tal que $P_{\mathcal{U}_n}(B_n - C_n) < \epsilon/2^n$. Sea $F_n = \pi_n^{-1}C_n$, entonces $F_n \subseteq E_n$ y $P(E_n - F_n) = P_{\mathcal{U}_n}(B_n - C_n) < \epsilon/2^n$. Como $\{F_n\}$ no tiene por qué ser una sucesión decreciente, tomamos $G_n = F_1 \cap \cdots \cap F_n$. Entonces tenemos que $G_n \subseteq F_n \subseteq E_n$ y además

$$G_n = \pi_1^{-1}(C_1) \cap \cdots \cap \pi_n^{-1}(C_n)$$

y como $\pi_i^{-1}(C_i) = \pi_n^{-1}(\pi_i^{-1}(C_i) \cap \mathcal{U}_n)$, tenemos que

$$G_n = \pi_n^{-1}((\pi_1^{-1}(C_1) \cap \cdots \cap \pi_{n-1}(C_{n-1}) \cap C_n) \cap \mathcal{U}_n)$$

Luego si ponemos $D_n = (\pi_1^{-1}(C_1) \cap \cdots \cap \pi_{n-1}(C_{n-1}) \cap C_n) \cap \mathcal{U}_n$, vemos que $G_n = \pi_n^{-1}(D_n)$ con D_n compacto. Además

$$P(E_n - G_n) = P(\bigcup_{i=1}^{n} E_n - F_i) \leq \sum_{i=1}^{n} P(E_n - F_i) \leq \sum_{i=1}^{n} P(E_i - F_i) < \epsilon$$

Pero $\{G_n\}$ decrece al vacío pues cada $G_n \subseteq E_n$ y por lo que probamos anteriormente $P(G_n) \downarrow 0$. Pero entonces

$$0 \leq \limsup_n P(E_n) \leq \lim_n P(G_n) + \epsilon = \epsilon$$

y como ϵ es arbitrario se concluye la demostración. $\square$

Teorema 3.4 (Minlos-Sazonov). *Sea $\varphi : H \to \mathbb{C}$ una función continua, definida positiva y tal que $\varphi(0) = 1$. Entonces φ es una función característica si y sólo si se verifica la siguiente condición: para todo $\epsilon > 0$ existe un S-operador en H, S_ϵ tal que si $\langle S_\epsilon x, x\rangle < 1$ entonces*

$$1 - \Re\varphi(x) < \epsilon$$

Demostración. Supongamos que φ es la transformada de Fourier de una medida de probabilidad P. Entonces claramente φ es continua, definida positiva y $\varphi(0) = 1$. Veamos que cumple la condición. Sea $\epsilon > 0$. Entonces

$$1 - \Re\varphi(x) = 2\int_H \operatorname{sen}^2\left(\frac{\langle y, x\rangle}{2}\right) dP(y)$$

$$= 2\int_{\|x\|\leq\delta} \operatorname{sen}^2\left(\frac{\langle y, x\rangle}{2}\right) dP(y) + 2\int_{\|x\|>\delta} \operatorname{sen}^2\left(\frac{\langle y, x\rangle}{2}\right) dP(y)$$

$$\leq \int_{\|x\|\leq\delta} \frac{\langle y, x\rangle^2}{2} dP(y) + 2P(\|x\| > \delta) < \int_{\|x\|\leq\delta} \frac{\langle y, x\rangle^2}{2} dP(y) + \frac{\epsilon}{2}$$

donde $\delta > 0$ lo elegimos para que $P(\|x\| > \delta) < \frac{\epsilon}{4}$. Definimos $\Lambda : H \times H \to \mathbb{R}$ como

$$\Lambda(v,w) = \int_{\|x\| \leq \delta} \langle y, v\rangle \langle y, w\rangle dP(y)$$

Entonces Λ es una forma bilineal simétrica en H y

$$|\Lambda(v,w)| \leq \delta^2 \|v\| \|w\|$$

Luego existe S operador lineal acotado en H tal que $\langle Sv, w\rangle = \Lambda(v,w)$ para todo $v, w \in H$. Se ve fácilmente que S es autoadjunto, positivo, compacto y de traza finita, es decir un S-operador en H. Sea $S_\epsilon = \frac{1}{\epsilon} S$ entonces tenemos que

$$1 - \Re\varphi(x) \leq \frac{\epsilon}{2} \langle S_\epsilon x, x\rangle + \frac{\epsilon}{2}$$

y esto nos da el operador que buscabamos.

Probemos el reciproco. Sea $\mathcal{U}$ un subespacio de dimensión finita de H. Entonces la restricción de φ a $\mathcal{U}$ es una función continua, definida positiva y tal que $\varphi(0) = 1$. Luego por el teorema de Bochner en dimensión finita sabemos que existe una medida de probabilidad $P_{\mathcal{U}}$ en $\mathcal{B}_{\mathcal{U}}$ tal que φ es su transformada de Fourier, es decir vale la siguiente represenatción para todo $h \in \mathcal{U}$,

$$\varphi(h) = \int_{\mathcal{U}} e^{i\langle x, h\rangle} dP_{\mathcal{U}}(x)$$

Probaremos que $\{P_{\mathcal{U}}\}$ es una familia compatible. Sean $\mathcal{U}_1 \subseteq \mathcal{U}_2$ subespacios de dimensión finita de H, entonces si $\pi : \mathcal{U}_2 \to \mathcal{U}_1$ denota la restricción de la proyección ortogonal sobre $\mathcal{U}_1$ a $\mathcal{U}_2$, lo que debemos probar es que $P_{\mathcal{U}_2}\pi^{-1} = P_{\mathcal{U}_1}$. Para ésto fijemos $x \in \mathcal{U}_1$. Si miramos a x como elemento de $\mathcal{U}_2$ podemos escribir

$$\int_{\mathcal{U}_1} \exp(i\langle x, y\rangle) dP_{\mathcal{U}_1}(y) = \varphi(x) = \int_{\mathcal{U}_2} \exp(i\langle x, y\rangle) dP_{\mathcal{U}_2}(y)$$

$$= \int_{\mathcal{U}_2} \exp(i\langle x, \pi(y)\rangle) dP_{\mathcal{U}_2}(y) = \int_{\mathcal{U}_1} \exp(i\langle x, y\rangle) dP_{\mathcal{U}_2}\pi^{-1}(y)$$

y la igualdad que buscabamos se deduce de la unicidad de la transformada de Fourier, en este caso para medidas en $\mathcal{U}_1$. Por la proposición anterior bastará ver que para cada $\epsilon > 0$ existe N_0 tal que si $N \geq N_0$ se tiene que

$$\sup P_{\mathcal{U}}(\{\|x\| \geq N\} \cap \mathcal{U}) < \epsilon$$

para probar la existencia de una medida de probabilidad P cuyas proyecciones sean $P_{\mathcal{U}}$. Veamos primero lo siguiente, dado $\alpha > 0$ y $N \geq 1$ tenemos que

$$\int_{\mathcal{U}} 1 - \exp\left(-\frac{\alpha}{2}\|x\|^2\right) dP_{\mathcal{U}}(x) \geq \int_{\|x\| \geq N} 1 - \exp\left(-\frac{\alpha}{2}\|x\|^2\right) dP_{\mathcal{U}}(x)$$

$$\geq \int_{\|x\|\geq N} 1-\exp\left(-\frac{\alpha}{2}N^2\right) dP_{\mathcal{U}}(x) = \left(1-\exp\left(-\frac{\alpha}{2}N^2\right)\right) P_{\mathcal{U}}(\|x\|\geq N)$$

Observemos lo siguiente, el integrando que utilizamos es la transformada de Fourier de una distribución normal multivariada en $\mathcal{U}$ con media cero y matriz de covarianza αI. Entonces si $d = \dim(\mathcal{U})$ escribiendo el integrando como la integral de $\cos\langle x,u\rangle z_\alpha(u) d\lambda_d(u)$ con $u \in \mathcal{U}$; donde hemos denotado z_α la densidad gaussiana y λ_d la medida de Lebesgue; vemos que podemos hacer Fubini e intercambiar las integrales para llegar a la siguiente igualdad:

$$\int_{\mathcal{U}} 1-\exp\left(-\frac{\alpha}{2}\|x\|^2\right) dP_{\mathcal{U}}(x) = \int_{\mathcal{U}} (1-\Re\varphi(u)) \frac{1}{(2\pi\alpha)^{d/2}} \exp\left\{-\frac{1}{2\alpha}\|u\|^2\right\} d\lambda_d(u)$$

$$\leq \int_{\mathcal{U}} \left(\frac{\epsilon}{2} + 2\langle S_\epsilon u, u\rangle\right) \frac{1}{(2\pi\alpha)^{d/2}} \exp\left\{-\frac{1}{2\alpha}\|u\|^2\right\} d\lambda_d(u)$$

donde S_ϵ es el S-operador que sabemos existe por hipótesis. Sea $\{e_i\}$ una base ortonormal completa de H tal que $\{e_1,\ldots,e_d\}$ es base de $\mathcal{U}$. Entonces como

$$\int_{\mathcal{U}} \langle u,e_i\rangle\langle u,e_j\rangle z_\alpha(u) d\lambda_d(u) = \delta_{ij}$$

tenemos que

$$\int_{\mathcal{U}} \left(\frac{\epsilon}{2} + 2\langle S_\epsilon u, u\rangle\right) \frac{1}{(2\pi\alpha)^{d/2}} \exp\left\{-\frac{1}{2\alpha}\|u\|^2\right\} d\lambda_d(u)$$

$$= \frac{\epsilon}{2} + 2\sum_{i,j=1}^{d} \langle S_\epsilon e_i, e_j\rangle \int_{\mathcal{U}} \langle u,e_i\rangle\langle u,e_j\rangle z_\alpha(u) d\lambda_d(u)$$

$$= \frac{\epsilon}{2} + 2\sum_{i=1}^{d} \langle S_\epsilon e_i, e_i\rangle \int_{\mathcal{U}} \langle u,e_i\rangle^2 \frac{1}{(2\pi\alpha)^{d/2}} \exp\left\{-\frac{1}{2\alpha}\|u\|^2\right\} d\lambda_d(u)$$

$$= \frac{\epsilon}{2} + 2\alpha\sum_{i=1}^{d} \langle S_\epsilon e_i, e_i\rangle \leq \frac{\epsilon}{2} + 2\alpha \mathrm{tr}(S_\epsilon)$$

En resumen, tenemos que

$$P_{\mathcal{U}}(\|x\|\geq N) \leq \left(\frac{\epsilon}{2} + 2\alpha\mathrm{tr}(S_\epsilon)\right)\left(1-\exp\left(-\frac{\alpha}{2}N^2\right)\right)^{-1}$$

y eligiendo α primero y N después podemos hacer dicha probabilidad arbitrariamente pequeña.

Hemos probado entonces que existe una medida de probabilidad P en H tal que sus proyecciones son $P_{\mathcal{U}}$. Veamos pues que φ es en efecto la transformada de Fourier de P. Sea $\mathcal{U}_n$

una sucesión creciente a un subespacio denso. Denotamos por P_n la medida de probabilidad en $\mathcal{U}_n$ y π_n la proyección. Entonces, dado $x \in H$ tenemos que

$$\varphi(\pi_n(x)) = \int_{\mathcal{U}_n} \exp(i\langle \pi_n(x), y\rangle) dP_n(y) = \int_H \exp(i\langle \pi_n(x), \pi_n(y)\rangle) dP(y)$$

$$= \int_H \exp(i\langle \pi_n(x), y\rangle) dP(y)$$

Como $\pi_n(x) \to x$ para todo $x \in H$ el teorema de convergencia dominada y la continuidad de φ nos conducen a la igualdad

$$\varphi(x) = \int_H \exp(i\langle x, y\rangle) dP(y)$$

Es decir, φ es la transformada de Fourier de P. □

3.4. Distribución gaussiana y el teorema central del límite.

En esta sección trabajaremos en el conjunto $\mathcal{P}_2$ de las medidas de probabilidad en un espacio de Hilbert separable H, tales que

$$\int_H \|x\|^2 \, dP(x) < \infty$$

Si $P \in \mathcal{P}_2$ vemos que $\int_H \|x\| \, dP(x) < \infty$ también y por la desigualdad de Cauchy-Schwarz también lo es

$$\int_H \langle x, y\rangle dP(x) < \infty$$

para todo $y \in H$. Además como la aplicación $y \mapsto \int_H \langle x, y\rangle dP$ es lineal y continua existe $u_P \in H$ unívocamente determinado por P tal que

$$\langle u_P, y\rangle = \int_H \langle x, y\rangle dP(x) \quad \forall y \in H$$

Por esta última igualdad vemos que u_P no es más que el valor esperado de un elemento aleatorio en H que tenga a P como distribución.

De forma totalmente análoga vemos que

$$\int_H \langle x, y\rangle\langle x, z\rangle dP(x) < \infty$$

para toda $P \in \mathcal{P}_2$ y todo $y, z \in H$. Como la aplicación

$$(y, z) \mapsto \int_H \langle x, y\rangle\langle x, z\rangle dP(x) - \langle u_P, y\rangle\langle u_P, z\rangle$$

es bilineal y continua existe un operador lineal acotado $S_P : H \to H$ determinado de manera única por la distribución P, que cumple

$$\langle S_P y, z\rangle = \int_H \langle x, y\rangle\langle x, z\rangle dP(x) - \langle u_P, y\rangle\langle u_P, z\rangle = \int_H \langle x - u_P, y\rangle\langle x - u_P, z\rangle dP(x)$$

Así en particular vemos que $\langle S_P y, y\rangle = \int_H \langle x - u_P, y\rangle^2 dP \geq 0$ para todo $y \in H$. Llamaremos a este operador, el operador de covarianza de la distribución P. Resumimos en la siguiente proposición las propiedades de S_P.

Proposición 3.4. *Si $P \in \mathcal{P}_2$ y S_P es el operador de covarianza de P, entonces*

(i) S_P es autoadjunto.

(ii) S_P es positivo, esto es $\langle S_P y, y\rangle \geq 0$ para todo $y \in H$.

(iii) S_P es un operador compacto, esto es la imagen de un conjunto acotado es relativamente compacta.

(iv) S_P tiene traza finita.

Demostración. (i) Basta observar que la forma bilineal que define S_P es simétrica, esto es $\omega(y, z) = \omega(z, y)$.

(ii) Ya lo probamos arriba.

(iii) y (iv) Para probar que S_P es compacto bastará probar que se cumple (iv). Sea $\{e_i\}$ una base ortonormal completa de H. Entonces como

$$\langle S_P e_i, e_i\rangle = \int_H \langle x, e_i\rangle^2 dP(x) - \langle u_P, e_i\rangle^2$$

para cada n tenemos que

$$\sum_{i=1}^{n} \langle S_P e_i, e_i\rangle = \int_H \sum_{i=1}^{n} \langle x, e_i\rangle^2 dP(x) - \sum_{i=1}^{n} \langle u_P, e_i\rangle^2$$

Y como el integrando de la derecha esta acotado por $\|x\|^2$ que tiene integral finita vemos que podemos pasar al límite y entonces

$$tr(S_P) = \sum_{i=1}^{\infty} \langle S_P e_i, e_i\rangle = \int_H \sum_{i=1}^{\infty} \langle x, e_i\rangle^2 dP(x) - \sum_{i=1}^{\infty} \langle u_P, e_i\rangle^2$$

$$= \int_H \|x - u_P\|^2 dP(x) < \infty$$

Como S_P es postivo existe el operador raíz cuadrada $S_P^{1/2}$ y tenemos entonces que

$$\sum_i \left\| S_P^{1/2} e_i \right\|^2 = \sum_i \langle S_P^{1/2} e_i, S_P^{1/2} e_i \rangle = \sum_i \langle S_P e_i, e_i \rangle = tr(S_P) < \infty$$

Por lo que la raíz de S_P es un operador de Hilbert-Schmidt y en consecuencia es compacto. Como los operadores compactos son un ideal en los operadores de H y $(S_P^{1/2})^2 = S_P$ concluimos que S_P es compacto.

□

A todo operador lineal acotado que cumpla las condiciones (i),(ii),(iii) y (iv) lo llamaremos un S-operador.

Proposición 3.5. *Sea $\varphi : H \to \mathbb{C}$ la función definida por*

$$\varphi(y) = \exp\left(i\langle \mu, y \rangle - \frac{1}{2}\langle Sy, y \rangle \right)$$

donde $\mu \in H$ y S es un S-operador. Entonces existe una medida de probabilidad P en H tal que $\varphi = \hat{P}$. Además $P \in \mathcal{P}_2$, μ es el valor medio de P y S es el operador de covarianza de P.

Demostración. Claramente φ es continua y $\varphi(0) = 1$. Sea $\{e_i\}$ una base ortonormal de H y denotemos V_n el subespacio de dimensión finita generado por $\{e_1, \ldots, e_n\}$. Entonces si $x = \sum_{i=1}^n x_i e_i \in V_n$ tenemos que

$$\varphi(x) = \exp\left(i \sum_{i=1}^n x_i \langle \mu, e_i \rangle + \frac{1}{2} \sum_{i,j=1}^n x_i x_j \langle S e_i, e_j \rangle \right)$$

Esto muestra que φ restringida a V_n es la transformada de Fourier de una distribución normal en $\mathbb{R}^n$ con media $(\langle \mu, e_i \rangle)_{i=1}^n$ y matriz de covarianza $\Sigma^n = (\langle S e_i, e_j \rangle)_{i,j=1}^n$ -recordar que S es autoadjunto y positivo-. Luego cada restricción a los V_n debe ser una función definida positiva. Llamemos φ_n a φ restringida a V_n. Sean $z_1, \ldots, z_N$ números complejos y $h_1, \ldots, h_N$ elementos de H. Para cada n ponemos $h_i^n = \pi_n(h_i) = \sum_{j=1}^n \langle h_i, e_j \rangle e_j \in V_n$ para $i = 1, \ldots, N$, donde π_n es la proyección ortogonal sobre V_n. Así tenemos que

$$\sum_{k,l=1}^N z_k \overline{z}_l \varphi(h_k^n - h_l^n) = \sum_{k,l=1}^N z_k \overline{z}_l \varphi_n(h_k^n - h_l^n) \geq 0$$

Cuando $n \to \infty$, $\pi_n(h_i) \to h_i$ y como φ es continua resulta que $\varphi(h_k^n - h_l^n) \to \varphi(h_k - h_l)$ por lo que tomando límite

$$\sum_{k,l=1}^N z_k \overline{z}_l \varphi(h_k - h_l) \geq 0$$

Es decir, φ es definida positiva. Para terminar la prueba por el teorema de Minlos-Sazonov, será suficiente mostrar que para cada $\epsilon > 0$ existe un S-operador S_ϵ tal que para todo $x \in H$ que virifique $\langle S_\epsilon x, x\rangle < 1$ se tiene $0 \leq 1 - \Re\varphi(x) < \epsilon$.

Por un lado

$$0 \leq 1 - \Re\varphi(x) = 1 - \cos\langle\mu, x\rangle \exp\left(-\frac{1}{2}\langle Sx, x\rangle\right)$$

$$= 1 - \exp\left(-\frac{1}{2}\langle Sx, x\rangle\right) + (1 - \cos\langle\mu, x\rangle)\exp\left(-\frac{1}{2}\langle Sx, x\rangle\right)$$

$$\leq \frac{1}{2}\langle Sx, x\rangle + \frac{1}{2}\langle\mu, x\rangle^2 = \langle\left(\frac{S + T_\mu}{2}\right)x, x\rangle = \epsilon\langle S_\epsilon x, x\rangle$$

donde $T_\mu x = \langle\mu, x\rangle\mu$ y $S_\epsilon = (S+T_\mu)/2\epsilon$. Luego bastará probar que S_ϵ es un S-operador. Como $\langle T_\mu x, x\rangle = \langle\mu, x\rangle^2 \geq 0$ para todo $x \in H$ tenemos que T_μ es un operador positivo. Además es compacto pues su imagen tiene dimensión uno. Se ve fácilmente que T_μ es autoadjunto y además

$$tr(T_\mu) = \sum_i \langle T_\mu e_i, e_i\rangle = \|\mu\|^2 < \infty$$

Por lo que T_μ es de hecho un S-operador. Luego al ser S_ϵ combinación lineal positiva de S-operadores también es un S-operador.

Probemos ahora que la medida de probabilidad P definida por φ está efectivamente en $\mathcal{P}_2$. No perdemos generalidad con suponer que $\mu = 0$. Entonces

$$\Re\varphi(x) = \exp\left(\frac{1}{2}\langle Sx, x\rangle\right) = \int_H \cos\langle y, x\rangle dP(y) = 1 - 2\int_H \operatorname{sen}^2\frac{\langle y, x\rangle}{2} dP(y)$$

$$\leq \exp\left(-2\int_H \operatorname{sen}^2\frac{\langle y, x\rangle}{2} dP(y)\right)$$

de donde deducimos que

$$2\int_H \operatorname{sen}^2\frac{\langle y, x\rangle}{2} dP(y) \leq \frac{1}{2}\langle Sx, x\rangle$$

para todo $x \in H$. Evaluando esta última desigualdad en tx con $t \in \mathbb{R}$ positivo y tomando límite en $t \to 0$ tenemos que

$$\liminf_{t\to 0} \frac{2}{t^2}\int_H \operatorname{sen}^2\frac{t\langle y, x\rangle}{2} dP(y) \leq \frac{1}{2}\langle Sx, x\rangle \quad \forall x \in H$$

Como el integrando $t^{-2}\operatorname{sen}^2(t\langle y, x\rangle/2) \to \langle y, x\rangle^2/4$ cuando $t \to 0$, el lema de Fatou nos dice que

$$\int_H \langle y, x\rangle^2 dP(x) \leq \langle Sx, x\rangle \quad \forall x \in H$$

Luego deducimos que

$$\int_H \sum_{i=1}^{n} \langle y, e_i \rangle^2 dP(y) \leq \sum_{i=1}^{n} \langle Se_i, e_i \rangle \leq tr(S) < \infty$$

y por el teorema de convergencia dominada deducimos que

$$\int_H \|y\|^2 \, dP(y) \leq tr(S) < \infty$$

Probaremos ahora que μ y S son el valor medio y el operador de covarianza de P respectivamente. Para esto sea v un elemento aleatorio con distribución P. Consideremos $\langle \cdot, h \rangle \in H^*$ para un $h \in H$ fijo, entonces tenemos que $\langle v, h \rangle$ es una variable aleatoria real tal que su función característica esta dada por

$$\begin{aligned}
\varphi_{\langle v,h \rangle}(t) = \int_{\mathbb{R}} e^{its} dF(s) &= \int_H e^{it\langle x,h \rangle} dP(x) = \hat{P}(th) \\
&= \exp\left(it\langle \mu, h \rangle - \frac{t^2}{2} \langle Sh, h \rangle \right)
\end{aligned}$$

Luego para todo $h \in H$ tenemos que $\langle v, h \rangle$ es una variable aleatoria con distribución normal, $E\langle v, h \rangle = \langle \mu, h \rangle$ y $\sigma^2 (\langle v, h \rangle) = \langle Sh, h \rangle$. Esto implica por definición de valor medio que μ es el valor esperado de v. Además tenemos que

$$\sigma^2 (\langle v, h \rangle) = \int_{\mathbb{R}} (s - \langle \mu, h \rangle)^2 dF(s) = \int_H \langle x - \mu, h \rangle^2 dP(x) = \langle Sh, h \rangle$$

para todo $h \in H$ y por la identidad de polarización esto define de manera única a S y vemos que $S = S_P$.

□

Definición 3.2. *Diremos que un elemento aleatorio en H tiene distribución gaussiana o que es gaussiano, si su distribución P tiene como transformada de Fourier*

$$\hat{P}(h) = \exp\left(i\langle \mu, h \rangle - \frac{\langle Sh, h \rangle}{2} \right)$$

donde en esta igualdad $\mu \in H$ y S es un S-operador.

Proposición 3.6. *Sea v un elemento aleatorio en H. Entonces v tiene distribución gaussiana si y sólo si $\langle v, h \rangle$ es una variable aleatoria con distribución normal para todo $h \in H$.*

Demostración. En la demostración del teorema anterior vimos que si un elemento aleatorio es gaussiano, entonces para todo $h \in H$, se tiene que $\langle v, h\rangle$ tiene distribución normal. Es decir, nos falta probar el recíproco.

Supongamos que $\langle v, h\rangle$ es normal para todo $h \in H$ de media μ_h y varianza σ_h^2. Denotamos por P la distribución de v. Entonces pongamos $v = \sum_i v_i e_i$ donde $\{e_i\}$ es una base ortonormal de H y $v_i = \langle v, e_i\rangle$. Por hipótesis cada v_i es una variable aleatoria normal de parámetros μ_i y σ_i^2. Además tenemos que

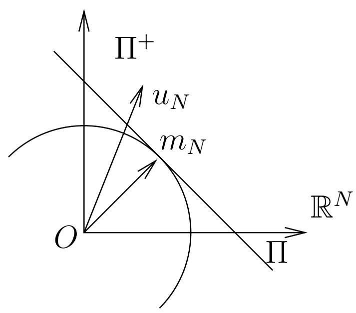
$$\sum_i v_i^2 = \|v\|^2 = \rho^2$$

por lo que la serie de la izquierda converge c.s. a una variable aleatoria finita ρ^2. Para cada N consideremos el vector aleatorio en $\mathbb{R}^N$ dado por $u_N = (v_1, \ldots, v_N)$. Luego u_N tiene distribución gaussiana. En efecto dado $(t_1, \ldots, t_N) \in \mathbb{R}^N$ arbitrario tenemos que

$$t_1 v_1 + \cdots + t_N v_N = \langle v, \sum_{i=1}^{N} t_i e_i\rangle$$

que tiene distribución gaussiana. El vector de coordenadas $m_N = (\mu_1, \ldots, \mu_N)$ no es otra cosa que el valor medio de dicho vector aleatorio. Además tenemos que

$$\rho_N^2 = \|u_N\|^2 = \sum_{i=1}^{N} v_i^2 \leq \rho^2$$



Sea Π el “plano” perpendicular -en $\mathbb{R}^N$- al subespacio generado por m_N -es decir a la recta $\overline{Om_n}$- que pasa por el punto m_N. Vemos que la probabilidad de que u_n este en uno u otro semi-espacio es la misma, es decir $P\left(u_N \in \Pi^+\right) = 1/2$, por lo que con probabilidad mayor o igual que $1/2$ tenemos que $\|m_N\| \leq \rho$.Esto implica que si la sucesión

$$\|m_N\|^2 = \sum_{i=1}^{N} \mu_i^2$$

no estuviera acotada, tendríamos una probabilidad superior o igual a 1/2 de que ρ^2 tampoco lo estuviera lo cual es imposible. Luego necesariamente debe ser

$$\sum_i \mu_i^2 < \infty$$

Podemos definir entonces el vector $\mu \in H$ poniendo $\mu = \sum_i \mu_i e_i$. Como sabemos que v tiene esperanza si y sólo si $v - \mu$ la tiene, podemos suponer que $\mu = 0$.

Consideremos $\pi_N : H \to V_N$ las proyecciones ortogonales sobre los subespacios V_N generados por $\{e_1, \ldots, e_N\}$. Sea $h \in H$ fijo, como $\pi_N(x) \to x$ para todo x,

$$\mathrm{E}\exp(it\langle \pi_N(v), h\rangle) = \int_H \exp(it\langle \pi_N(x), h\rangle) dP(x)$$

$$= \int_H \exp(it\langle x, \pi_N(h)\rangle dP(x) = \varphi_v(\pi_N(th)) \to \varphi_v(th) = \mathrm{E}\exp(it\langle v, h\rangle)$$

para todo $t \in \mathbb{R}$. Luego la función característica φ_N de $\langle \pi_N(v), h\rangle$ converge puntualmente a φ la función característica de $\langle v, h\rangle$. Además como son variables gaussianas, existen los momentos y podemos escribir

$$\varphi(t) = 1 + i\mu_h t + r_1(t) \text{ y } \varphi_N(t) = 1 + r_2(t)$$

donde

$$\left|\frac{r_i(t)}{t}\right| \to 0 \text{ si } |t| \to 0$$

Luego fijemos $t > 0$, entonces

$$|\mu_h||t| \leq |\varphi(t) - \varphi_N(t)| + |r_1(t) - r_2(t)|$$

y haciendo $N \to \infty$ tenemos que

$$|\mu_h| \leq \frac{|r_1(t)| + |r_2(t)|}{|t|} \to 0$$

cuando $|t| \to 0$. Esto muestra que $\mathrm{E}\langle v, h\rangle = 0$ para todo $h \in H$, de donde existe $\mathrm{E}v = 0$. En general -cuando $\mu \neq 0$- se tiene $\mathrm{E}v = \mu$.

De manera muy similar, utilizando el mismo tipo de razonamientos pero de una forma un tanto más engorrosa, se puede probar que $\mathrm{E}\|v\|^2 < \infty$.[2] Esto implica que $P \in \mathcal{P}_2$ y nos

[2]El lector interesado puede consultar E. Mourier, Annales de l'Inst. H. Poincaré, vol. 13, no. 3, 1953, pág. 161-244.

garantiza la existencia del S-operador de covarianza S de P. Por definición tenemos entonces que

$$\langle Sh, h\rangle = \sigma^2\left(\langle v, h\rangle\right) = \sigma_h^2$$

Luego para cada $t \in \mathbb{R}$ y $h \in H$ fijos vale que

$$\varphi_v(th) = \mathrm{E}\exp(it\langle v, h\rangle) = \exp\left(it\mu_h - \frac{t^2}{2}\sigma_h^2\right) = \exp\left(it\langle \mu, h\rangle - \frac{t^2}{2}\langle Sh, h\rangle\right)$$

Esto vale en particular para $t = 1$, lo que muestra que v es un e.a. gaussiano.

□

La proposición anterior es la que permite conectar la definición que nosotros dimos de e.a. gaussiano en un espacio de Hilbert, con su equivalente para espacios de Banach. Nuestro último teorema es una generalización del teorema central del límite para v.a. i.i.d. en $\mathbb{R}$.

Teorema 3.5. *Sea $\{v_n\}$ una sucesión de elementos aleatorios independientes e identicamente distribuídos en H, tal que la distribución de v_1 está en $\mathcal{P}_2$ y $E(v_1) = 0$. Si S es el operador de covarianza de v_1 y definimos*

$$z_n = \frac{1}{n^{1/2}}\sum_{i=1}^{n} v_i$$

Entonces

$$z_n \Rightarrow N$$

donde N es la distribución gaussiana con media 0 y operador de covarianza S.

Demostración. Sea $h \in H$ fijo. Entonces $\{\langle v_n, h\rangle\}$ es una sucesión de variables aleatorias independientes e identicamente distribuídas en $\mathbb{R}$ con media cero y varianza $\langle Sh, h\rangle$. Luego por el teorema central de límite en $\mathbb{R}$, sabemos que

$$\langle z_n, h\rangle = \frac{1}{n^{1/2}}\sum_{i=1}^{n}\langle v_i, h\rangle \Rightarrow N(0, \langle Sh, h\rangle)$$

Como la transformada de Fourier de $\langle z_n, h\rangle$ es

$$\varphi_{\langle z_n, h\rangle}(t) = \hat{P_{z_n}}(th)$$

y como

$$\varphi_{\langle z_n, h\rangle}(t) \rightarrow_n \exp\left(\frac{-t^2}{2}\langle Sh, h\rangle\right)$$

para todo $t \in \mathbb{R}$, tenemos que en particular para $t = 1$,

$$\hat{P_{z_n}}(h) \to_n \exp\left(\frac{-\langle Sh, h\rangle}{2}\right)$$

Como esto vale para todo $h \in H$ tenemos la convergencia puntual. Veamos ahora que $\{P_{z_n}\}$ es tensa.

Fijemos $\{e_i\}$ una base ortonormal completa de H. Entonces para cada $m \geq N$,

$$\int_H \sum_{i=N}^{m} \langle x, e_i\rangle^2 dP_{z_n}(x) = \sum_{i=N}^{m} \int_H \langle x, e_i\rangle^2 dP_{z_n}(x) = \sum_{i=N}^{m} \sigma^2\left(\langle z_n, e_i\rangle\right) = \sum_{i=N}^{m} \langle Se_i, e_i\rangle$$

y como podemos pasar al límite pues la varianza es finita, tenemos que

$$\int_H r_N^2(x) dP_{z_n}(x) = \sum_{i=N}^{\infty} \langle Se_i, e_i\rangle$$

para todo n. Luego como esta serie converge pues es la traza de S, tenemos que el límite en N es cero. Entonces la sucesión es tensa. El teorema de continuidad nos garantiza que P_{z_n} converge en distribución a una medida de probabilidad cuya función carcaterística es (*). $\square$

Bibliografía

[*] Para los capítulos 1 y 2, la bibliografía utilizada fue esencialmente Padgett-Taylor[4] y Revesz[6]. Para el capitulo 3 utilizamos Billingsley[1], Grenander[2], Laha-Rohatgi[3] y Parthasarathy[5].

[1] P. Billingsley - Convergence of Probability Measures. John Wiley & Sons. New York, 1968.

[2] U. Grenander - Probabilities on Algebraic Structures. Stockholm, New York, Almqvist & WiksellWiley, 1963.

[3] R.G. Laha and V.K. Rohatgi - Probability Theory. John Wiley & Sons. 1979.

[4] W.J. Padgett and R.L. Taylor - Laws of Large Number for Normed Linear Spaces and Certain Fréchet Spaces. Lecture Notes in Mathematics. Springer-Verlag 1973.

[5] K.R. Parthasarathy - Probability Measures on Metric Spaces. New York, Academic Press. 1967.

[6] P. Revesz - The laws of large numbers.- Academic Press, New York, 1968.