UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

"TOPOLOGÍA Y PROBABILIDAD"

T E S I S

QUE PARA OBTENER EL GRADO ACADÉMICO DE

MATEMÁTICO

P R E S E N T A

ADRIÁN GONZÁLEZ CASANOVA SOBERÓN

DIRECTORA DE TESIS: DRA. MARÍA EMILIA CABALLERO

México, D.F. NOVIEMBRE 2009.

Hoja de Datos del Jurado

1. Datos del alumno Apellido paterno Apellido materno Nombre(s) Teléfono Universidad Nacional Autónoma de México Facultad de Ciencias Carrera Número de cuenta	1. Datos del alumno González Casanova Soberón Adrián 56558188 Universidad Nacional Autonóma de México Facultad de Ciencias Matemáticas 3-0278604-8
2. Datos del tutor Grado Nombre(s) Apellido paterno Apellido materno	2. Datos del tutor Dra. María Emilia Caballero Acosta
3. Datos del sinodal 1 Grado Nombre(s) Apellido paterno Apellido materno	3. Datos del sindal 1 Matemático José Luis Pérez Garmendia
4. Datos del sinodal 2 Grado Nombre(s) Apellido paterno Apellido materno	4. Datos del sinodal 2 Dr. Victor Manuel Rivero Mercado
5. Datos del sinodal 3 Grado Nombre(s) Apellido paterno Apellido materno	5. Datos del sinodal 3 Dr. Juan Carlos Pardo Millan
6. Datos del sinodal 4 Grado Nombre(s) Apellido paterno Apellido materno	6. Datos del sinodal 4 Dr. Gerónimo Francisco Uribe Bravo
7. Datos del trabajo escrito Título Subtitulo Número de páginas Año	7. Datos del trabajo escrito Topología y Probabilidad Espacios métricos de utilidad probabilistica 121 2009

Índice general

Introducción

La convergencia es uno de los conceptos cuya noción varía más radicalmente durante el estudio de una carrera de matemáticas. Desde los primeros semestres, el alumno es sorprendido al comprender conceptos como límites y sucesiones de números reales, que rápidamente se vuelven más complejos. Cuando empieza a sentirse cómodo con estas ideas, finalmente descubre que la convergencia es algo mas general, algo que caracteriza la topología de un espacio. El eje sobre el que giran todos los capítulos de esta tesis es conocido como convergencia débil, la convergencia débil es la convergencia que surge del teorema del limite central y de hecho resulta sumamente útil en todas las ramas de la probabilidad. Este concepto está íntimamente vinculado con una métrica: la métrica de Prohorov definidas en el conjunto de todas las medidas de probabilidad en $(S, B(S))$, donde S es un espacio metrico. Un resultado muy importante es que la convergencia débil se puede metrizar y la métrica que lo logra es la de Prohorov.

Una afirmación que puede servir como motivación es la siguiente, a la que llamaremos pregunta 1: Los Procesos Poisson Compuestos son densos en los procesos de Lévy con la topología inducida por la convergencia débil. ¿ Que significa esto? Aún suponiendo que el lector está familiarizado con lo que es un proceso estocástico (una coleccion de variables aleatorias), es probable que no le quede claro que significa que 2 procesos converjan y por lo tanto que no comprenda el significado de que un subconjunto de los procesos de Lévy sea denso. Todos estos temas se tocarán con rigor mas adelante, pero ahora mismo diremos que la manera más fácil de concebir la convergencia de procesos estocasticos es pensando en estos como medidas de probabilidad sobre el espacio de las trayectorias (trayectorias continuas por la derecha con límite por la izquierda (CADLAG) en particular).

Se verá que una sucesión de medidas converge débilmente a otra si y sólo si converge en la métrica de Prohorov (bajo hipótesis comunes sobre el espacio métrico (S, d) sobre el que estén definidas las medidas). Es por ello que este trabajo se inicia con un capítulo dedicado al estudio de la métrica de Prohorov sobre un espacio métrico (S, d). En este capítulo estudiamos las propiedades del espacio, así como la manera de calular la distancia entre dos medidas de probabilidad. Para entender el espacio de Prohorov hacemos una digresión que nos ayuda a entender el espacio en general, estudiamos la distancia de Prohorov de medidas sobre el intervalo real $[0, 1]$. Las conclusiones de esta digresión nos permiten describir los compactos y asegurar que sobre el intervalo se pueden conocer la distancia explícita entre cualesquiera 2 medidas finitas.

El espacio métrico subyacente determina muchas de las propiedades del espacio de Prohorov, por ejemplo si (S, d) es los reales con la distancia euclidiana se tienen medidas de probabilidad sobre $\mathbb{R}$ o, lo que es equivalente, variables aleatorias con valores en los reales y se trata del concepto tradicional de convergencia débil o en distribución. Si (S, d) es $(c[a, b], \sup)$ el espacio de prohorov es el de los procesos estocasticos continuos. Si (S, d) es $(D[0, T], \cdot)$ el espacio de prohorov es el de los procesos estocastocos. Como lo que nos interesa son los procesos estocasticos en general es preciso estudiar el espacio de las CADLAG, como un espacio métrico, dicho espacio es conocido como el espacio de Skorohod y es bastante mal portado, como se mostrará. Dotarlo con una metrica que haga que las cosas que queremos que converjan en efecto converjan no es trivial (por ejemplo la métrica uniforme de las funciones continuas, no hace que las indicadoras de $[1 - \frac{1}{n}), 2)$ convergan a la indicadora de $[1, 2)$, al contrario, siempre distan 1. En el capítulo 2 realizamos muchos ejemplos explícitos de distancia según las métricas propuestas por Skorohod para que el lector se familiarice con este espacio que además de útil es muy bonito. En este capítulo también estudiaremos a la convergencia débil en función de la convergencia finito dimensional y la tensión.

Para entender la pregunta 1 nos falta entender el concepto de proceso de Lévy y porqué es relevante. En el capítulo tercero explicamos los principales resultados de procesos de Lévy y los ejemplares mas útiles de estos, como lo son el Movimiento Browniano y el proceso Poisson compuesto.

En el último capitulo mezclamos todos los ingredientes para probar algunos resultados de convergencia en procesos de Lévy, probamos que para asegurar la convergencia de una sucesión de procesos de Lévy a otro proceso de Lévy, es suficiente probar la convergencia a tiempo 1, lo cual es asombroso. Por último damos respuesta formal a la pregunta 1 y exhibimos la relación del espacio métrico de los procesos de Lévy con otros espacios métricos de interés.

Estas hojas están pensadas para que el lector logre entender y sentirse comodo con temas que son muy útiles tanto en probabilidad teórica como aplicada, su rico contenido de ejemplos permite que el lector pueda captar conceptos que a simple vista son difíciles de descifrar. No se realizan todas las pruebas, pues el objetivo de este trabajo es que sea lo mas accecible y agradable que se pueda, los temas que estudiaremos están llenos de sutilezas, por lo que en algunas secciones se sacrificara un poco de formalidad para procurar que el lector no se agobie, que al contrario, se sorprenda.

Capítulo 1

Espacio de Prohorov

[1] ¿Qué tanto se parece una medida de probabilidad a otra? Esta pregunta es fácil de responder en algunos casos, por ejemplo cuando comparamos 2 normales es suficiente fijarnos en su esperanza y su varianza para darnos una idea de cuan parecidas son. Sin embargo, responder a esto en general es más complicado, le debemos a Prohorov la manera de comparar 2 medidas de probabilidad cualesquiera.

Estudiaremos al conjunto de las medidas de probabilidad sobre un espacio métrico S con métrica d y la sigma-algebra $\Im$ (que normalmente sera el conjunto de los Borelianos, denotado $B(S)$); se le dará a este conjunto una métrica, la métrica de Prohorov, para tener un nuevo espacio. La importancia de esta métrica radica en su relación con la convergencia débil, cuando una sucesión de medidas se acerca tanto como queramos a una medida, en términos de Prohorov, dicha sucesión converge también débilmente, profundizaremos en esto más adelante. (esta relación nos permite decir que la métrica de Prohorov, de cierta forma, cuantifica lo lejos que están dos medidas de que converjan débilmente).

Las medidas de probabilidad tienen una cercana relación con las variables aleatorias, una variable aleatoria X definida en $(\omega, \Im, P)$ con valores en $(S, B(S))$ induce una medida de probabilidad de la siguiente manera

$$P_X(A) = P(X \in A) \quad \text{con} \quad A \in B(S)$$

inversamente, para toda medida de probabilidad m en $(\omega, \Im)$, existe una variable aleatoria, en un espacio de probabilidad $(\omega, \Im, P)$, con valores en $(S, B(S))$ tal que

$$P(X_m \in A) = m(A) \quad \text{con} \quad A \in B(S)\,.$$

[1](Ver [9], [2] y [3])

Así, todos los resultados de esta sección pueden verse en el sentido de medidas o en el de variables aleatorias, en particular podemos definir la distancia de Prohorov en ambos casos. Vamos a definir algunas cosas antes de la métrica de Prohorov. Sean P, Q medidas de probabilidad en (S, d), y sea $\hbar$ la colección de subconjuntos cerrados de S. Si $F \in \hbar$, entonces definimos el ϵ-ensanchamiento de F como

$$F^{\epsilon} = \{x \in S \,:\, \inf_{y \in F} d(x, y) < \epsilon\} = \{x \in S \,:\, d(x, F) < \epsilon\},$$

que es el conjunto F más los puntos que están a distancia menor que ϵ de él. Otra manera de llamarlo es una ϵ-*vecindad* de F.

Algunas propiedades de las ϵ-*vecindad*, con estas definiciones son:

1. $F \subset F^{\epsilon}$

   Es claro. Para todo punto en F la distancia entre el y F es cero lo que es menor que ϵ.

2. F^{ϵ} es abierto

   Los puntos ahí cumplen que su distancia a F es estrictamente menor que ϵ de donde existe un real δ tal que $d(F, x) + \delta < \epsilon$, así que si tomamos una bola con centro en x y radio δ, por la desigualdad del triángulo, estará totalmente contenida en F^{ϵ} , i. e. todo punto es punto interior.

3. si G=S-F^{ϵ} entonces G es cerrado y $F \subset S - G^{\epsilon}$

   Basta notar que F^{ϵ} es abierto y que si $x \in F$, entonces $d(x, G) > \epsilon$ y por lo tanto $x \in S - G^{\epsilon}$

4. si $\delta, \epsilon > 0$ entonces $(\bar{F^{\delta}})^{\epsilon} \subset F^{\epsilon+\delta}$

   Basta tomar un $x \in \left(\bar{F^{\delta}}\right)^{\epsilon}$ y un $z \in \bar{F^{\delta}}$ tal que $d(z, x) < \epsilon$, entonces existe un $\bar{z} \in F^{\delta}$ tal que d$(\bar{z}, x) < \epsilon$, como $\bar{z} \in F^{\delta}$, existe $y \in F$ tal que d$(\bar{z}, y) < \delta$ la desiguldad del triángulo nos asegura que $d(y, x) < \epsilon + \delta$ i.e. $x \in F^{\epsilon+\delta}$.

5. sí a_n es una sucesión que decrece a cero y F cerrado, entonces $F = \cap_{n=1}^{\infty} F^{a_n}$.

   Sea $x \in F$ es claro que $x \in F^{a_n}$ pues $d(x, F) = 0$ i.e. $F \subset \cap_{n=1}^{\infty} F^{a_n}$.

   inversamente supongamos que $\cap_{n=1}^{\infty} F^{a_n} \subset F$ es una afirmación falsa, entonces existe $x \in \cap_{n=1}^{\infty} F^{a_n}$ tal que $x \in S - F$. Entonces existe $z_n \in F$

tal que $d(z_n, x) < a_n$ pero esto implica que z_n converge a x, lo cual contradice que F es cerrado. La contradicción viene de suponer que $\cap_{n=1}^{\infty} F^{a_n} \subset F$ es una afirmación falsa, lo que termina la prueba.

Ahora, dadas P, Q dos medidas de Probabilidad en (S, d) ó $(S, B(S))$, definimos la métrica de Prohorov de la siguiente manera:

$$\rho(P, Q) = \text{ínf}\left\{\epsilon > 0 : P(F) \leq Q(F^{\epsilon}) + \epsilon, \forall F \in \hbar\right\}.$$

Nos acostumbraremos a esta definición más adelante en los ejemplos. Ahora demostremos que en efecto es métrica, pero antes necesitamos un lema

Lema 1.1. *Si P, Q son medidas de probabilidad sobre S y $\epsilon > 0$ y se cumple que*

$$P(F) \leq Q(F^{\epsilon}) + \epsilon \quad \textit{para todo} \quad F \in \hbar$$

entonces

$$Q(F) \leq P(F^{\epsilon}) + \epsilon \quad \textit{para todo} \quad F \in \hbar$$

Demostración. Supongamos que

$$P(F) \leq Q(F^{\epsilon}) + \epsilon \quad \text{para todo} \quad F \in \hbar\,,$$

definimos $G = S - F^{\epsilon}$, Por las propiedades de ϵ-vesindad 1 y 3 se obtiene que

$$P(F^{\epsilon}) = 1 - P(G) \geq 1 - Q(G^{\epsilon}) - \epsilon = Q(S - G^{\epsilon}) - \epsilon \geq Q(F) - \epsilon$$

es decir

$$Q(F) \leq P(F^{\epsilon}) + \epsilon\,,$$

esto termina la prueba. □

Ahora sí, podemos proceder a probar que es métrica

- Primero probaremos que $\rho(P, Q) = \rho(Q, P)$.

  Basta mostrar que $\{\epsilon : P(F) \leq Q(F^{\epsilon}) + \epsilon \text{ para todo } F \in \hbar\} = \{\epsilon : Q(F) \leq P(F^{\epsilon}) + \epsilon \text{ para todo } F \in \hbar\}$, pues esto implica que sus ínfimos son el mismo. Sin embargo, por el lema anterior, si $\epsilon \in \{\epsilon : P(F) \leq Q(F^{\epsilon}) + \epsilon \text{ para todo } F \in \hbar\}$ también está en $\{\epsilon : Q(F) \leq P(F^{\epsilon}) + \epsilon \text{ para todo } F \in \hbar\}$, lo que prueba la simetría.

- $\rho(P,Q) = 0$ si solo si $P = Q$.

  Supongamos que $\rho(P,Q) = 0$, entonces $P(F) \leq Q(F^{\frac{1}{n}}) + \frac{1}{n}$ para todo $F \in \hbar$, para todo $n \in N$. Por la propiedad 5 y tomando límites obtenemos que

  $$P(F) \leq \lim_{n\to\infty} Q(F^{\frac{1}{n}}) + \frac{1}{n} = Q(\cap_1^\infty F^{\frac{1}{n}}) = Q(F)\,.$$

  Por lo tanto

  $$P(F) \leq Q(F), \quad \text{para todo} \quad F \in \hbar$$

  pero la simetría implica que

  $$P(F) = Q(F), \quad \text{para todo} \quad F \in \hbar\,.$$

  La igualdad sobre todos los conjuntos cerrados implica la igualdad de las medidas, por argumentos estandar de teoría de la medida y el hecho de que los cerrados generan a los Borelianos.

  Recíprocamente, si $P = Q$ es claro que $\rho(P,Q) = 0$ pues el conjunto $\{\epsilon > 0 : P(F) \leq Q(F^\epsilon) + \epsilon\}$ consiste en todos las $\epsilon > 0$, por las propiedades 1 y 5, de donde el ínfimo es cero.

- $\rho(P,Q) \leq \rho(P,R) + \rho(R,Q)$

  Supongamos que $\rho(P,R) = \delta$ y $\rho(R,Q) = \gamma$, sea $\delta' > \delta$ y $\gamma' > \gamma$ entonces

  $$P(F) \leq R(F^{\delta'})+\delta' \leq R(\overline{F^{\delta'}})+\delta' \leq Q(\overline{F^{\delta'}}^{\gamma'})+\gamma'+\delta' \leq Q(F^{\delta,+\gamma'})+\gamma'+\delta'$$

  Por la propiedad 4. Entonces $\rho(P,R) \leq \delta' + \gamma'$ para todo $\delta' > \delta$ y $\gamma' > \gamma$ lo que implica que $\rho(P,R) \leq \delta + \gamma$.

Esto implica que es métrica, y mas aún, se trata de una métrica acotada por 1, pues sabemos que toda medida de probabilidad es acotada por 1, de donde la ecuación

$$P(F) \leq Q(F^1) + 1\,,$$

siempre es válida.

Esta métrica no es intuitiva, y en general no queda claro cómo calcular la distancia de una medida a otra, para familiarizarnos con este extraño espacio estudiaremos algunos casos en los que sí se puede saber la distancia exacta entre dos probabilidades, así como también veremos que hay algunas cotas fáciles de calcular para la distancia de cualquier pareja de medidas.

Lo primero que podemos decir es el siguiente lema

Lema 1.2. *Sean P,Q medidas de probabilidad. Si existe un cerrado F y una $\epsilon > 0$ tales que $P(F) > 0$ y $Q(F^\epsilon) = 0$, entonces $\rho(P,Q) > 0$*

Basta usar la defininción de la métrica de Prohorov para notar que, de hecho, $\rho(P,Q) \geq \text{mín}(P(F), \epsilon) > 0$.

Ahora tomemos un ejemplo que debería ser fácil para observar los problemas técnicos en el cálculo de la distancia de Prohorov y veamos que podemos decir.

Ejemplo 1.3. Sea U_1, U_2 las medidas asociadas a variables aleatorias uniformes en los intervalos (x_1, y_1) y (x_2, y_2) respectivamente, supongamos que $(x_2, y_2) \subset (x_1, y_1)$, veamos que se puede decir de la distancia entre estas medidas en el sentido de Prohorov.

Como $(x_2, y_2) \subset (x_1, y_1)$, tenemos que $|x_1 - y_1| \geq |x_2 - y_2|$, entonces hay 2 regiones de relevancia el intervalo (x_2, y_2), al que llamaremos R_1 en donde la medida U_2 asigna valores mas grandes o iguales a los que asigna U_1, y la región R_2, $(x_1, x_2) \bigcup (y_2, y_1)$ en la que U_2 es idénticamente cero y U_1 asigna valores positivos, llamaremos R_3 al complemento de la unión de los intervalos.

Lema 1.4.

$$\rho(U_1, U_2) \leq U_2(R_1) - U_1(R_1) = 1 - U_1(R_1)$$

Demostración. Sea F un cerrado, se puede escribir como la unión de $G_i = R_i \cap F$ con $i \in \{1, 2, 3\}$, entonces $U_2(F) = \bigcup_1^3 U_2(G_i) = U_2(G_1)$ notemos que $U_2(G_1) > U_1(G_1)$ mientras que $U_2(G_2) \leq U_1(G_2)$, entonces si A es un cerrado que está en R_1, $U_2(A) - U_1(A) \geq 0$ como esto es para cualquier A, y por aditividad de las medidas $U_2(R_1) - U_1(R_1) \geq U_2(A) - U_1(A)$, pero esto no es sólo para A en R_1, es para cualquier A, pues afuera de la región R_1 la diferencia de las medidas es negativa, es decir si B esta afuera de R_1, $U_2(B) - U_1(B) = -U_1(B) \leq 0$, y por sigma aditividad $U_2(R_1) - U_1(R_1) \geq U_2(F) - U_1(F)$ para todo conjunto F. Esto nos da una cota muy simple de computar, tenemos que $U_1(F) \leq U_1(F^\epsilon)$, entonces $U_2(F) - U_1(F^\epsilon) \leq U_2(F) - U_1(F) \leq U_2(R_1) - U_1(R_1) = 1 - U_1(R_1)$, lo que termina la prueba. □

Este razonamiento nos sera útil durante los siguientes capitulos, mas adelante formalizaremos la definición de conjunto de máxima diferencia, sin embargo la estimación anterior no siempre es cercana a la distancia real, en particular si tomamos una uniforme $(0, 1)$ y una $(0, 1 + \epsilon)$ la cota que acabamos de sacar nos dice $\rho(U_1, U_{1+\epsilon}) \leq \epsilon$ sin embargo tenemos el siguiente lema

Lema 1.5. *Bajo las hipótesis anteriores*

$$\rho(U_1, U_{1+\epsilon}) = \frac{\epsilon}{2+\epsilon}\,,$$

El conjunto que maximiza $U_1(F) - U_{1+\epsilon}(F^\delta)$ es el intervalo $(0,1)$, pues si nos fijamos en un intervalo mas grande la diferencia disminuye ya que la parte negativa aumenta, mientras que si nos fijamos en uno más pequeño, quitaríamos algo positivo, por lo que nos alejamos del máximo. Ahora pensemos $\delta < \epsilon$, queremos encontrar

$$\rho(U_1, U_{1+\epsilon}) = inf\{\delta : U_1(F) - U_{1+\epsilon}(F^\delta) \leq \delta, \forall F \in \hbar\}\,,$$

como $(0,1)$ es el intervalo que maximiza diferencia queremos conocer el ínfimo de las δ que cumplen

$$U_1((0,1)) - U_{1+\epsilon}((0,1)^\delta) \leq \delta\,,$$

tenemos que

$$U_1((0,1)) - U_{1+\epsilon}((0,1)^\delta) = 1 - \frac{1+\delta}{1+\epsilon} = \frac{\epsilon-\delta}{1+\epsilon}\,,$$

entonces ese ínfimo se alcanza cuando $\delta = \frac{\epsilon-\delta}{1+\epsilon}$, lo que pasa si sólo si $\delta = \frac{\epsilon}{2+\epsilon}$, por lo tanto:

$$\rho(U_1, U_{1+\epsilon}) \leq \frac{\epsilon}{2+\epsilon}\,,$$

Para probar la igualdad basta notar que cualquier valor $k < \frac{\epsilon}{2+\epsilon}$ cumple que $U_1((0,1)) - U_{1+\epsilon}((0,1)^k) > k$ lo que se deduce de que la función $f(x) = 1 - \frac{1+x}{1+\epsilon} - x = \frac{\epsilon-(2+\epsilon)x}{1+\epsilon}$ es decreciente. Entonces

$$\rho(U_1, U_{1+\epsilon}) \leq \frac{\epsilon}{2+\epsilon}\,,$$

lo cual es un poco menos de la mitad de la cota que habíamos calculado antes.

1.1 La distancia para medidas de Dirac

Ahora nos enfocaremos en las medidas de Dirac, es decir,las que valen 1 si un punto x está en el conjunto, cero si no está. Estas medidas son de interés porque, como mostraremos más adelante, las combinaciones convexas de ellas forman un conjunto denso de medidas, así que entender cómo se portan es acercarnos a entender la geometría en el espacio de Prohorov.

Pensemos en una medida M de probabilidad qué acumula toda la masa en x, estudiaremos qué medidas distan de esta menos de ϵ, en el sentido de Prohorov.

Sea $B_{(\epsilon,x)}$ una bola de radio ϵ alrededor de x, en el sentido de la distancia en el espacio S.

Teorema 1.6. *Sea Q medida de probabilidad y $M = \delta_x$ una medida de Dirac, entonces:*

- $Q(B_{(\epsilon,x)}) \geq 1 - \epsilon$ *implica que* $\rho(Q, M) \leq \epsilon$.
- $\rho(Q, M) < \epsilon$ *implica que* $Q(B_{(\epsilon,x)}) \geq 1 - \epsilon$.

Demostración. $A_1 \subset B_{(\epsilon,x)}$ implica que $x \in A_1^\epsilon$, entonces

$$Q(A_1) \leq M(A_1^\epsilon) = 1\,,$$

si ahora tomamos $A_2 \subset B^c_{(\epsilon,x)}$, por ser Q medida tenemos que

$$Q(A_2) \leq Q(B^c_{(\epsilon,x)}) \leq \epsilon\,.$$

Sea F un conjunto cerrado arbitrario, podemos ver a F como $F = A_1 \bigcup A_2$ donde $A_1 = F \cap B_{(\epsilon,x)}$ y $A_2 = F \cap B^c_{(\epsilon,x)}$, de donde usando las 2 ecuaciones de arriba concluimos que

$$Q(F) = Q(A_1) + Q(A_2) \leq M(A_1^\epsilon) + \epsilon = M(F^\epsilon) + \epsilon$$

de donde $\rho(M, Q) \leq \epsilon$.

Ahora veamos que esta es una propiedad necesaria. Supongamos que $Q(B_{(\epsilon,x)}) < 1 - \epsilon$, entonces existe un conjunto cerrado F contenido en $B^c_{(\epsilon,x)}$ tal que $Q(F) > \epsilon$, claramente $M(F^\epsilon) = 0$ pues $d(F, x) > \epsilon$, entonces $Q(F) > M(F^\epsilon) + \epsilon$ por lo que Q no está en la bola de radio ϵ y centro en x en el sentido de la métrica de Prohorov.

Esto nos dice que si una medida concentra en una bola de radio ϵ, alrededor de x toda la masa excepto a lo más ϵ la distancia de Prohorov de esta medida y $M = \delta_x$ es menor que ϵ, además la distancia exacta entre una medida y δ_x es el ínfimo número positivo tal que esto se cumple. □

De esto concluimos el siguiente corolario

Corolario 1.7. *La distancia entre dos medidas de Dirac, es el mínimo entre la distancia de los puntos en que acumulan la masa y* 1.

Esto es claro pues si decimos que $M = \delta_x$ y $Q = \delta_y$, toda bola alrededor de x cuyo radio sea mayor a la distancia entre x y y cumplirá la propiedad necesaria y suficiente que expusimos, mientras que si la distancia es menor a 1 toda bola de radio menor a dicha cantidad no la cumplirá, lo que asegura que la distancia es el ínfimo, si es mayor a 1, sabemos que la métrica esta acotada por 1.

Ahora analizemos la distancia entre $M = \delta_y$ y $Q = a\delta_{x_1} + (1-a)\delta_{x_2}$, usando lo que hemos probado, para saber cual es la distancia entre M y Q basta saber, cual es el conjunto de las ϵ tales que la medida Q asigna una probabilidad menor a ϵ al complemento de la bola de radio ϵ alrededor de y. Sea δ la distancia entre el más cercano a y de los puntos que acumulan la masa segun la medida Q y Δ la distancia del más lejano,

$$\delta = \text{mín}\{d(x_1, y), d(x_2, y)\}$$
$$\Delta = \text{máx}\{d(x_1, y), d(x_2, y)\}$$

llamemos

$$\kappa = \text{ínf}(k \geq \delta \;\; \text{tal que} \;\; k \geq a_\Delta)\,,$$

donde a_Δ es la masa que acumula Q en el punto más alejado de y tenemos el siguiente resultado:

Corolario 1.8. *La distancia entre $M = \delta_y$ y $Q = a\delta_{x_1} + (1-a)\delta_{x_2}$ en términos de Prohorov es:*

$$\rho(M, Q) = \text{mín}(1, \Delta, \kappa)\,.$$

Demostración. Primero supongamos que κ es el mínimo y que $\kappa < \Delta$, se cumple que para todo $\kappa' > \kappa$, que $Q(B_{(\kappa', y_1)}) = 1 - a_\Delta \geq 1 - \kappa'$, y para todo $k < \kappa$ esto no se cumple, por lo tanto κ es la distancia de Prohorov, por el primer teorema de la sección.

Si Δ es el mínimo, la bola de radio Δ alrededor de y es la más pequeña con la propiedad que expusimos, ya que claramente esta bola tiene medida mayor que $1 - \epsilon$ (en particular tiene medida 1) y como es el mínimo no hay otra mas pequeña.

Por ultimo la prueba es trivial si 1 es el mínimo. $\square$

Para pasar de aquí a una medida que acumule la masa en N puntos haremos un razonamiento similar, sea $Q = \sum_1^N a_i \delta_{x_i}$, M igual que anteriormente, claramente la distancia está acotada superiormente por uno y por Δ (donde Δ es la distancia al más lejano), si la distancia es menor que estas dos cotas el razonamiento para calcularla es igual que el anterior. Sin embargo ahora la k no solo tiene que ser mayor o igual que el más lejano coeficiente a_Δ, tiene que ser mayor que la suma de todos los coeficientes cuyo punto base, x_i tiene una distancia mayor que k. Sin perdida de generalidad diremos que x_1 es el más cercano a y y que están ordenados hasta llegar al más lejano que es x_N, llamaremos d_i a la distancia entre y y x_i, entonces tenemos el siguiente teorema

Corolario 1.9. *la distancia entre M y Q en términos de la métrica de Prohorov es*

$$\rho(M,Q) \leq \text{mín}(1, \Delta, \text{ínf}(k \geq d_1 : k \geq \sum_{n_0}^{N} a_i))$$

donde $n_0 = \text{mín}\{m : d(y, x_m) \geq k\}$.

Pudimos ver quién está en la bola de radio ϵ alrededor de una medida que acumula la masa en un punto, pues notemos que al analizar el caso $Q = \sum_1^N a_i \delta_{x_i}$ por densidad de estas medidas, también estamos acomodando a todas las continuas, el siguiente paso para entender la geometría del espacio de Prohorov sería entender quién está en la bola de radio ϵ de una medida de la forma $Q = \sum_1^N a_i \delta_{x_i}$, de nuevo por densidad de estas medidas, esto nos permitiría conocer la geometría de todo el espacio, sabríamos quién está junto a quién y nos permitiría computar la distancia entre 2 medidas arbitrarias.

1.2 Aproximadores

En esta sección mostraremos que es posible aproximar la distancia, al menos en $\mathbb{R}$, en el sentido de Prohorov, entre cualesquiera dos medidas de probabilidad, esto no es algo que se ocurra simple, pues considerar el conjunto de todos los cerrados de un conjunto, en ningun caso es tarea fácil y en algunas situaciones ni siquiera es posible.

Lo que haremos es estudiar una familia de medidas de probabilidad, suficientemente simple como para que podamos calcular la distancia entre elementos de esta y suficientemente grande para que sea denso.

Trabajaremos con medidas en el intervalo cerrado $[0,1] \subset \mathbb{R}$, sin embargo, los resultados que aquí obtengamos se generalizan trivialmente a cualquier intervalo finito, y atraves de truncación, usando el hecho de que para toda medida de probabilidad P sobre $\mathbb{R}$, y para todo $\epsilon > 0$, existe un intervalo cerrado y finito F, tal que

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

se puede extender los resultados a medidas de probabilidad sobre los reales. Pasar a $\mathbb{R}^n$ tambien es sencillo si se piensa a este como un espacio producto.

Ahora sí, definamos a las medidas que nos interesan. Sea μ el conjunto de las medidas de probabilidad sobre el intervalo $[0,1]$. Sean $x_i = \frac{i}{2^n}$ con $i \in \{0, 1, 2, \ldots, 2^n\}$. Las medidas que estudiaremos son el conjunto.

$$\varsigma_n = \{m \in \mu : m(\{x_0, x_1, x_2 ... x_{2^n}\}) = 1\}$$

si $m \in \varsigma_n$ podemos escribir a m como

$$m = \sum_{i=0}^{2^n} \alpha_i \delta_{x_i} .$$

Definimos el conjunto

$$\varsigma = \bigcup_{n \in N} \varsigma_n$$

probemos que ς es denso.

Lema 1.10. *ς es denso.*

Demostración. Sea P una medida de probabilidad sobre $[0,1]$, fijemos n, entonces tenemos las siguientes igualdades

$$1 = P([0,1]) = P(\bigcup_{i=1}^{2^n} (x_{i-1}, x_i] \bigcup \{0\}) = P(\{0\}) + \sum_{i=1}^{2^n} P((x_{i-1}, x_i]) ,$$

sea $\alpha_0 = P(0)$ y $\alpha_i = P((x_{i-1}, x_i])$ para $i \in \{1, 2.., 2^n\}$. Definimos

$$m_n = \sum_{i=0}^{2^n} \alpha_i \delta_{x_i}$$

claramente $m_n \in \varsigma_n$, ademas notemos que si F es un cerrado de $[0,1]$

$$m_n(F) = \sum_{x_i \in F} \alpha_i = \sum_{x_i \in F} P((x_{i-1}, x_i]) \leq P(F^{\frac{1}{2^n}}) ;$$

donde la última desigualdad es porque si $x_i \in F$ se cumple que $(x_{i-1}, x_i] \subset F^{\frac{1}{2^n}}$. Pero el hecho de que $m_n(F) \leq P(F^{\frac{1}{2^n}})$ implica que

$$\rho(m_n, P) < \frac{1}{2^n}$$

si definimos analogamente m_n para todo n, tenemos que

$$m_n \to P \quad \text{en el sentido de Prohorov}$$

pero como claramente $\{m_n\}_{n\geq 1} \subset \varsigma$, lo que nos asegura que, en efecto ς es denso. $\square$

Definición 1.11. Diremos que $m_n \in \varsigma$ es un aproximador de P, si

$$m_n = \sum_{i=0}^{2^n} \alpha_i \delta_{x_i}$$

donde $\alpha_0 = P(0)$ y $\alpha_i = P((x_{i-1}, x_i])$ para $i \in \{1, 2, \ldots, 2^n\}$.

El siguiente paso sera exhibir que la distancia entre 2 elementos de ς_n es computable, esto nos interesa pues, si queremos calcular la distancia entre Q, P medidas de probabilidad sobre $[0,1]$, podemos tomar m_n y v_n en ς, tales que m_n es aproximador de P y v_n aproximador de Q. La desigualdad del triángulo nos asegura que

$$\rho(P,Q) = \lim_{n\to\infty} \rho(m_n, v_n)$$

mas aún

$$\rho(m_n, v_n) - \frac{1}{2^{n-1}} \leq \rho(P,Q) \leq \rho(m_n, v_n) + \frac{1}{2^{n-1}}\,.$$

Entonces si pudiéramos computar la distancia entre cualesquiera dos elementos de ς_n esto implica que podemos aproximar tanto como queramos la distancia entre dos medidas de probabilidad arbitrarias.

Antes de mostrar que podemos calcular la distancia de dos elementos de la familia que estamos estudiando, necesitamos una definición

Definición 1.12. Dadas Q y P medidas de probabilidad, diremos que un conjunto F_0 es ϵ-conjunto de máxima diferencia si para todo cerrado F, se cumple que

$$P(F_0) - Q(F_0^{\epsilon}) \geq P(F) - Q(F^{\epsilon}) \;\text{ o }\; Q(F_0) - P(F_0^{\epsilon}) \geq Q(F) - P(F^{\epsilon})\,.$$

Conocer un conjunto de máxima diferencia nos es útil pues si F_0 es ϵ-conjunto de máxima diferencia y $P(F_0) - Q(F_0^{\epsilon}) = k$, es claro que si $k \leq \epsilon$

$$\rho(P,Q) = inf\{\delta : P(F) - Q(F^{\delta}) \geq \delta, paratodoF \in \hbar\} \leq \epsilon\,.$$

Un caso de interés es si nos fijamos en 2 medidas P, Q sobre los enteros, en ese caso, el ensanchamiento de un punto con masa nunca alcanza otro punto con masa, ya que la métrica de Prohorov esta acotada por 1, i.e.

$$P(\{x\}) = P(x^\epsilon) \ \text{ y } \ Q(\{x\}) = Q(x^\epsilon) \ \text{ para todo } x \text{ entero y } \epsilon \in (0,1]$$

esto nos permite asegurar que si un conjunto de puntos con masa, es ϵ de máxima diferencia para alguna ϵ, lo es para toda ϵ. De donde $\rho(P,Q)$ es simplemente $P(F_0) - Q(F_0^\epsilon)$ donde F_0 es un conjunto de máxima diferencia compuesto por enteros. Es fácil ver que el conjunto de máxima diferencia que nos funciona es

$$F_0 = \{x \in E : P(x) < Q(x)\} \ \text{ o } \ F_0 = \{x \in E : P(x) > Q(x)\}$$

de donde concluimos que calcular la distancia de 2 medidas de probabilidad sobre los enteros es simple.

Lema 1.13. *La distancia entre cualesquiera Q, M elementos de ς_n es computable.*

Demostración. Sea $\epsilon > 0$, sea χ el conjunto potencia de $\{x_1, x_2, \ldots, x_{2^n}\}$, sea F un cerrado, afirmamos que existe un conjunto C en χ tal que

$$P(C) - Q(C^\epsilon) \geq P(F) - Q(F^\epsilon)\,,$$

para probar esto definimos $C = \{x_1, x_2, \ldots, x_{2^n}\} \cap F$, claramente esta en χ, además $P(C) = P(F)$, mientras que $Q(C^\epsilon) \leq Q(F^\epsilon)$, lo que nos asegura la afirmación.

Entonces es inteligente construir el conjunto de máxima diferencia como un subconjunto de χ.

Queremos construir $F_0 \in \chi$ tal que

$$P(F_0) - Q(F_0^\epsilon) \geq P(C) - Q(C^\epsilon)\,,$$

para todo $C \in \chi$.

Si

$$a_0 = \text{máx}\{P(F) - Q(F^\epsilon,) \colon F \in \chi\}\,,$$

entonces es ϵ-conjunto de máxima diferencia todo $F_0 \in \chi$ tal que

$$a_0 = P(F_0) - Q(F_0^\epsilon)\,.$$

Podemos afirmar que para todo $\epsilon > 0$ existe un ϵ-conjunto de máxima diferencia pues el conjunto χ tiene cardinalidad finita.

Ya exhibimos que para toda ϵ existe un ϵ-conjunto de máxima diferencia, pero podemos decir más, notemos que si $C \in \chi$ para toda pareja $\epsilon_1, \epsilon_2 \in (\frac{i-1}{2^n}, \frac{i}{2^n}]$, tenemos que

$$Q(C^{\epsilon_1}) = Q(C^{\epsilon_2}) \text{ y } P(C^{\epsilon_1}) = P(C^{\epsilon_2})\,.$$

Por lo tanto si F_0 es ϵ_1-conjunto de máxima diferencia, también es ϵ_2-conjunto de máxima diferencia. Esto nos asegura que existen solo 2^n conjuntos de máxima diferencia, a los que denotaremos $F_1, F_2, \ldots, F_{2^n}$, si $\epsilon \in (\frac{i-1}{2^n}, \frac{i}{2^n}]$.

Sea $k_i = P(F_i) - Q(F_i^{\frac{i}{2^n}})$, una observación importante es que la sucesión k_i es no creciente, pues $F_i^{\frac{j}{2^n}}$ es creciente como función de j y $P(F_i)$ es constante. Con todo esto estamos listos para calcular la distancia de Prohorov de Q y M, definimos $I = \text{mín}(i : k_i < \frac{i}{2^n})$, entonces

$$\rho(M, Q) = \text{máx}(k_I, \frac{I-1}{2^n})\,.$$

Probar esto es simple, primero veamos que $\pi = \text{máx}(k_I, \frac{I-1}{2^n})$ cumple que

$$P(F) - Q(F^{\pi+\delta}) \leq \pi + \delta \text{ para todo } F \text{ cerrado, para } \delta > 0\,.$$

Sea F un cerrado, supongamos que $\delta < \frac{1}{2^n}$, si delta es más grande la prueba es aun mas simple, por lo que la dejamos al lector, entonces

$$P(F) - Q(F^{\pi}) \leq P(F_I) - Q(F^{\frac{I}{2^n}}) = k_I \leq \pi \leq \pi + \delta\,.$$

Ahora supongamos que existe $\epsilon < \pi$ tal que

$$P(F) - Q(F^{\epsilon}) \leq \epsilon \text{ para todo } F \text{ cerrado}$$

claramente para alguna i, $\epsilon \in [\frac{i-1}{2^n}, \frac{i}{2^n})$, entonces en particular para F_i el ϵ-conjunto de máxima diferencia, se cumple que

$$P(F_i) - Q(F_i^{\epsilon}) = k_i \leq \epsilon < \frac{i}{2^n}\,.$$

Luego, $i = I$ i.e. $k_i = k_I$, entonces $k_I \leq \epsilon$, si $\pi = k_I$ ya acabamos, de no ser así, al suponer que $\epsilon < \pi$, esto nos asegura que

$$k_I \leq \epsilon < \frac{I-1}{2^n} = \frac{i-1}{2^n}$$

lo cual es una contradicción dado que $\epsilon \in [\frac{i-1}{2^n}, \frac{i}{2^n})$, lo que concluye la prueba. $\square$

1.3 El ρ-intervalo

Usaremos lo estudiado en el capitulo de Aproximadores para probar que el conjunto de las medidas con soporte contenido en el intervalo unitario es compato. Trabajaremos sobre los reales.

En este capitulo, el ρ-intervalo, I, se definine como el conjunto de las medidas de probabilidad con soporte contenido en el intervalo unitario en los reales i.e. $I = \{m : m([0,1]) = 1\}$.

Lema 1.14. *El ρ-intervalo es cerrado.*

Demostración. Sea m una medida de probabilidad tal que su soporte no esta contenido en el intervalo unitario de los reales, es decir $m \in I^c$. Entonces existe un cerrado F tal que $m(F) > 0$ y $F \cap [0,1] = \emptyset$, por regularidad de las medidas en los reales, además como $[0,1]$ es compacto la distancia de F a $[0,1]$ es positiva, esto implica, por el lema 2 de este capitulo, que existe δ, tal que para toda $v \in I$, $\rho(m,v) \geq \delta$, lo que implica que no existe una sucesión de medidas contenidas en I que converga a m, es decir m no es punto de acumulación de I. Como esto fue para toda m en I^c, todos los puntos de acumulación de I estan en I y por lo tanto I es cerrado. □

Para probar que I es totalmente acotado usaremos el siguiente lema

Lema 1.15. *El conjunto de las medidas con soporte en $D_n = \{x_1, x_2, \ldots, x_{2^n}\}$ $= \{\frac{i}{2^n} : i \in \{1, 2, \ldots, 2^n\}\}$ y masa total menor o igual que 1, al que denotaremos L_{2^n}, es totalmente acotado, para toda n.*

Demostración. Sea n fija. Sea L_k El conjunto de las medidas con soporte en $\{x_1, x_2, \ldots, x_k\}$
$= \{\frac{i}{2^n} : i \in \{1, 2, \ldots, k\}\} \subset D_n$ y masa total menor o igual que 1. Lo que probaremos es que para todo $t > n$, el conjunto

$$\omega_t = \{m \in L_{2^n} : m = \sum_{i \in D_n} \frac{j_i}{2^t} \delta_{x_i}\}$$

cumple que

$$L_{2^n} \subset \bigcup_{m \in \omega_t} B_{(m, \frac{1}{2^t})} .$$

Vamos a usar indución de la siguiente forma, primero probaremos que para todo $t > n$

$$\omega_t^1 = \{m \in L_{2^n} : m = \frac{j}{2^t} \delta_{x_1}\}$$

cumple que

$$L_1 \subset \bigcup_{m \in \omega_t^1} B_{(m, \frac{1}{2^t})}.$$

Luego veremos que es verdad para dos puntos, es decir, para todo $t > n$

$$\omega_t^2 = \{m \in L_{2^n} : m = \frac{j_1}{2^t}\delta_{x_1} + \frac{j_2}{2^t}\delta_{x_2}\}$$

cumple que

$$L_2 \subset \bigcup_{m \in \omega_t^2} B_{(m, \frac{1}{2^t})}.$$

Usaremos que es verdad para dos puntos para probar que también lo es para tres, es decir, para todo $t > n$

$$\omega_t^3 = \{m \in L_{2^n} : m = \frac{j_1}{2^t}\delta_{x_1} + \frac{j_2}{2^t}\delta_{x_2}\frac{j_3}{2^t}\delta_{x_3}\}$$

cumple que

$$L_3 \subset \bigcup_{m \in \omega_t^3} B_{(m, \frac{1}{2^t})}.$$

Como la técnica para pasar de dos puntos a tres es análoga a la de pasar de k a $k+1$ con esto concluiremos la prueba. (Obs: Las bolas de las que hablamos son en el sentido de la métrica de Prohorov).

- El primer paso es trivial, ya que

  $$\rho(a\delta_x, b\delta_x) = |a - b|.$$

- Para el segundo paso tomamos $p \in L_2$. Definimos $w = a_1\delta_{x_1} + a_2\delta_{x_2}$ donde $a_1 = \frac{h_1}{2^t}$ y $a_2 = \frac{h_2}{2^t}$ donde $\frac{h_1}{2^t} \leq p(\{x_1\}) < \frac{h_1+1}{2^t}$ y $\frac{h_2-1}{2^t} \leq p(\{x_2\}) < \frac{h_2}{2^t}$.

  Claramente, como $\{x_1\}$ es conjunto de máxima diferencia,

  $$\rho(p, w) \leq \frac{1}{2^t}$$

  además, w esta en ω_t^2 lo que prueba el resultado para $k = 2$.

- Por último tomamos $q \in L_3$. Sea $a_3 = \frac{h_3}{2^t}$ donde $\frac{h_3}{2^{t+1}} \leq q(\{x_1\}) < \frac{h_3+1}{2^{t+1}}$. Sea $p = q1_{(\{x_1\},\{x_2\})}$, claramente $p \in L_2$, entonces existe $w \in \omega_2$ tal que $\rho(p, w) \leq \frac{1}{2^{t+1}}$, sea F_p conjunto de máxima diferencia de p y

w, tal que $0 < p(F_p) - w(F_p) < \frac{1}{2^{t+1}}$. Notemos que, por construcción, $\varpi = p + a_3\delta_{x_3}$ esta en ω_t^3 y el conjunto $F_p \bigcup\{x_3\}$ es conjunto de máxima diferencia, entonces

$$\rho(q,\varpi) \leq q(F_p \bigcup\{x_3\}) - \varpi(F_p \bigcup\{x_3\}) = q(F_p) - \varpi(F_p) + q(\{x_3\}) - \varpi(\{x_3\}) \leq \frac{1}{2^t}$$

lo que termina la prueba para $k = 3$, y como el caso k a $k+1$ es analogo terminamos la prueba.

□

Despues de probar que el lema anterior tenemos los siguientes 3 corolarios.

Corolario 1.16. *El conjunto de las medidas con soporte en* $\{\frac{i}{2^n} : i \in \{1, 2, \ldots, 2^n\}\}$ *y masa total menor o igual que* 1 *es compacto para toda* n.

Demostración. Basta notar que es cerrado, lo cual se demuestra igual que el primer lema de la sección. □

Corolario 1.17. *El conjunto de las medidas de probabilidad con soporte en* $\{\frac{i}{2^n} : i \in \{1, 2, \ldots, 2^n\}\}$ *es compacto para toda* n.

Demostración. Como esta contenido en el compacto anterior, basta ver que es cerrado, para esto es suficiente notar que una medida con masa total menor que uno tiene una distancia positiva de cualquier medida de probabilidad, esto se ve tomando $[0, 1]$ como el cerrado que cumple las hipotesis del lema 2 de este capítulo.

Demostración.

Corolario 1.18. *El intervalo es totalmente acotado.*

Demostración. Sea $n > 0$. Sea $x \in I$ sea α_{2n} un aproximador de x, tal que $\rho(x, \alpha_{2n}) \leq 2^{-2n}$. Sea m en $\hat{w}_t^n$, (donde $\hat{w}_t^n$ se define como w_t^n pero sólo con medidas de probabilidad y cumple lo mismo por el corolario anterior), tal que $\rho(m, \alpha_{2n}) \leq 2^{-2n}$. La desigualdad del triángulo nos asegura que $\rho(m, x) \leq 2^{-n}$ y como $\hat{w}_t^n$ es finito, esto termina la prueba. □

Teorema 1.19. *el* $\rho-$*intervalo es compacto*

Demostración. Es cerrado y totalmente acotado. □

Corolario 1.20. *La cerradura de las bolas en el* $\rho-$*intervalo, es compacta.*

Demostración. Es cerrado contenido en compacto. □

Todo esto nos podría llevar a suponer que que el espacio de Prohorov es localmente compacto, sin embargo esto es incorrecto, en el espacio de Prohorov la cerradura de las bolas no es compacta como mostraremos a continuación.

Pensemos en $B_{(x,\epsilon)}$. Sea A un conjunto finito tal que $x(A) > 1 - \epsilon$. Sea $k = \sup(A)$. Sea $x_1 = \mathbf{1}_A x + \epsilon\delta_k$ y $x_i = \mathbf{1}_A x + \epsilon\delta_{k+i2\epsilon}$. Notemos que $x_i \in B_{(x,\epsilon)}$, $\forall i > 0$ sin embargo $\rho(x_i, x_j) = \epsilon$, por lo que si una cubierta de $B_{x,\epsilon}$ con bolas de radio $\frac{\epsilon}{3}$ no puede tener a x_i, x_j de donde tiene que tener una infinidad de elementos. De donde $B_{x,\epsilon}$ no es uniformemente acotado.

Cabe resaltar que usando las mismas tecnicas con las que probamos que el ρ-intervalo es compacto, se puede ver que toda ρ-celda, C, definida como

$$C_{a,b,c,d} = \{m : m([a,b]^c) = 0, m([a,b]) \in [c,d]\} \text{ para } a < b \text{ y } c < d\,.$$

es compacta. Esto nos da intuición al respecto de uno de los teoremas mas importantes de este capitulo, el teorema de Prohorov, que nos dice que un conjunto A de medidas es Localmente compacto si existe un compacto en su soporte al que todas las medidas de A le asingnan una masa cercana a 1.

1.4 Propiedades del espacio de Prohorov

Ahora que hemos logrado familiarizarnos un poco con la métrica de Prohorov estudiaremos las propiedades que hacen que el espacio de Prohorov sea interesante.

La métrica de Prohorov tiene la importante propiedad de que si (S, d) es un espacio métrico separable, el conjunto de las medidas de probabilidad con la métrica de Prohorov también lo es, y que si (S, d) además es completo, el espacio de las medidas de probabilidad también es completo.

Lema 1.21. *El espacio de Prohorov sobre* $([0, 1], d)$ *es separable.*

Demostración. Basta notar que el subconjunto de los aproximadores tiene un subconjunto denso numerable.

Sea m un aproximador con soporte en los diadicos $D_n = \{\frac{k}{2^n} : k = 1, 2..,2^n\}$. Sea m_j un aproximador racional sobre los mismos puntos, es decir, m_j es un aproximador que a todo conjunto le asigna un valor racional, tal que $0 \leq m_j(\{0\}) - m(\{0\}) < \frac{1}{j}$ y $0 \leq \sum_1^{2^n} m(\{x_i\}) - m_j(\{x_i\}) < \frac{1}{j}$. claramente $\rho(m, m_j) < \frac{1}{j}$ □

No demostraremos un hecho mas general, de que si (S,d) es completo, entonces el espacio de Prohorov sobre (S,d) también lo es y que si (S,d) es separable, entonces el espacio de Prohorov sobre (S,d) también lo es.[2].

En el caso $S=\mathbb{R}$, otro denso de interés es el de las funciones absolutamente continuas con respecto a la de Lebesgue, el hecho de que sea denso lo que nos dice es que, dada cualquier medida de probabilidad sobre el espacio $(\Re,d)$ podemos construir una sucesión de medidas absolutamente continuas respecto a la de Lebesque, tal que converga a esta en el espacio de Prohorov.

La prueba es la siguiente

Demostración. Sea M una medida de probabilidad y sea P_n una sucesión de medidas de la forma $G=\sum_{i=1}^{N} a_i\delta_{x_i}$ (como las que usamos para probar la separabilidad) que converja a M. lo que haremos es tomar un elemento de P_n y construir una sucesión de medidas absolutamente continuas con respecto a la de Lebesgue que converja a ese elemento.

Sea $P\in G$ construiremos una sucesión de medidas absolutamente continuas con respecto a la de Lebesgue $(C_j)_{j>0}$ que converja a P: la gráfica de la densidad del j-ésimo elemento de la sucesión C_j, es una serie de triángulos isósceles, cada uno con la base centrada en x_i. La base de cada uno mide $\frac{1}{j}$ y altura $(2j)(a_i)$. Con esa construcción, la medida de la bola de radio $\frac{1}{2j}$ alrededor de x_i es simplemente el área del triángulo, es decir:

$$C_j(B_{(x_i,\frac{1}{2j})})=[\frac{1}{j}][(2j)(a_i)]/2=a_i\,.$$

Esto nos asegura que nuestra construcción en efecto es una sucesión de medidas de probabilidad, además como ya sabemos calcular distancias según Prohorov, es claro que

$$\rho(P,C_j)<\frac{1}{j}\,.$$

Esto fue para cualquier $P\in G$. El hecho de que G es denso termina la prueba. $\square$

A continuación demostraremos el teorema de Prohorov, pero antes definimos tensión:

Definición 1.22 (Tensión)**.** Decimos que una medida de probabilidad P, es tensa, si para todo $\epsilon>0$ existe un compacto $K\subset S$ tal que $P(k)\geq 1-\epsilon$. Una familia de medidas de probabilidad A es tensa si para todo $\epsilon>0$ existe un compacto $K\subset S$ tal que $\inf_{P\in A}P(K)\geq 1-\epsilon$.

[2]Una demostración se encuentra en *Markov Processes* de Ethier y Kurtz.

Para familiarizarnos con la definición de tensión probaremos el siguiente teorema.

Teorema 1.23. *Sobre todo espacio metrico (S,d) completo y separable, cualquier conjunto finito $P_1, P_2, \ldots, P_N$ de medidas de probabilidad es tenso.*

Demostración. Sea $\epsilon > 0$, para cada P_i existe un compacto K_i tal que $P_i(K_i) > 1 - \epsilon$, al cual construimos de la siguiente forma: como S es separable existe una colección numerable de bolas de radio $\frac{1}{n}$, B_i tales que $S = \bigcup_1^\infty B_i$. Es claro que $\lim_{n\to\infty} P_i(\bigcup_1^n B_i) = 1$. De lo anterior concluimos que para todo $\epsilon > 0$ existe n_0 tal que $P_i(\bigcup_1^n B_i) > 1 - \epsilon$. Como S es completo y $\bigcup_1^n B_i$ es totalmente acotado, su cerradura es compacta. LLamemos $K_i = cerr(\bigcup_1^n B_i)$. Sea $K = \bigcup_1^N K_1$. Como la union finita de compactos es compacta k es compacto, el hecho de que $P(K) \geq P_i(K_i) > 1 - \epsilon$ para $i = 1, 2 ... N$ termina la prueba. $\square$

Teorema 1.24 (Teorema de Prohorov)**.** *Si (S, d) es completo y separable, y A es una familia de medidas de probabilidades, es equivalente:*

1. *A es tenso i.e. para todo $\epsilon > 0$ existe un compacto $K \subset S$ tal que $\inf_{P\in A} P(K^\epsilon) \geq 1 - \epsilon$*

2. *A es relativamente compacto, i.e. su cerradura es un compacto.*

Demostración. 1). $\Longrightarrow$ 2). Por hipótesis A es tenso lo que implica que para todo $\epsilon > 0$ existe un compacto $K \subset S$ tal que $\inf_{P\in A} P(K^\epsilon) \geq 1 - \epsilon$. Ya mencionamos que si (S, D) es completo, también el espacio de las medidas de probabilidad sobre S es completo, como A es un subconjunto de las medidas de probabilidad sobre S, su cerradura es completa. Tenemos el teorema de Ascoli que nos dice que un conjunto es completo y totalmente acotado si solo si es compacto. Así que falta probar que A es totalmente acotado, es decir, para todo $\delta > 0$ existe un conjunto finito N de medidas de probabilidad sobre S tal que $A \subset \bigcup_{P\in N} B_\delta(P)$.

Sea $0 < \epsilon < \frac{\delta}{2}$, sea K compacto tal que $\inf_{P\in A} P(K^\epsilon) \geq 1 - \epsilon$, como K es compacto existe un conjunto finito $\{x_1, \ldots, x_n\} \in S$ tal que

$$K^\epsilon \subset \bigcup_{i=1}^n B_\delta(x_i), \text{ con } x_i \in S\,.$$

Sea $m > n$, proponemos

$$N = \{P = \sum_{i=0}^n \frac{k_i}{m}\delta_{x_i} : 0 \leq k_i \leq m, \sum_{i=0}^n k_i = m\}\,.$$

sea $Q \in A$, sea $k_i = mQ(E_i)$ donde $E_i = B_i - \bigcup_{j=1}^{i-1} B_j$ y sea $k_0 = m - \sum k_i$, claramente hay una $P_0 \in N$ con dichas k_i, probemos que $\rho(P_0, Q) < \delta$:

Sea F cerrado en S

$$\begin{aligned} Q(F) &= Q(F \cap (\bigcup_{i=1}^{n} E_i)) + Q(F \cap (\bigcup_{i=1}^{n} E_i)^c) \leq Q(F \cap (\bigcup_{F \cap E_i \neq \emptyset} E_i)) + \epsilon \\ &\leq \sum_{E_i \cap F \neq \emptyset} \frac{Q(E_i)m}{m} + \epsilon \leq \sum_{E_i \cap F \neq \emptyset} \frac{k_i}{m} + 2\epsilon \\ &\leq P_0(F^{2\epsilon}) + 2\epsilon \end{aligned}$$

Como esto fue para un cerrado arbitrario $\rho(P_0, Q) < 2\epsilon < \delta$ de donde A es totalmente acotado, y por lo tanto A es relativamente compacto.

2). $\Longrightarrow$ 1). Sea $\epsilon > 0$. Como A es totalmente acotado, Para toda n natural, existe un conjunto finito N_n contenido en A, tal que $A \subset \bigcup_{P \in N_n} B_{\frac{\epsilon}{2^{n+1}}}(P)$. Como probamos que una colección finita de medidas siempre es tensa, podemos tomar compactos K_n tales que $P(K_n) \geq 1 - \frac{\epsilon}{2^{n+1}}$ para todo P en N_n. Sea $Q \in A$, sabemos que Q esta en una bola de radio $\frac{\epsilon}{2^{n+1}}$ alrededor de algún elemento de N_n para todo n. Llamaremos P_n a cada uno de estas medidas. Para cada P_n existe un compacto K_n tal que $P_n(K_n) \geq 1 - \frac{\epsilon}{2^{n+1}}$ y $\rho(Q, P_n) < \frac{\epsilon}{2^{n+1}}$ de donde

$$Q(K_n^{\epsilon/2^{n+1}}) \geq P_n(K_n) - \epsilon/2^{n+1} \geq 1 - \frac{\epsilon}{2^n}\,.$$

Esto es para todo n, y definimos K como la cerradura de $\cap_{i \geq 1} K_i^{\epsilon/2^{i+1}}$, K es cerrado y totalmente acotado (por que $\cap_{i \geq 1} K_i^{\epsilon/2^{i+1}} \subset K_1^{\epsilon/2^{i+1}}$), por lo tanto es compacto y cumple

$$Q(K) \geq 1 - \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} = 1 - \epsilon\,.$$

Lo que nos asegura que en efecto A es tenso. □

Corolario 1.25. *Las medidas con soporte en el intervalo [0,1] son un conjunto compacto.*

Demostración. Sabemos que son cerradas y tienen soporte en un compacto, por lo que el teorema de Prohorov asugura que es un conjunto compacto. □

Otro resultado interesante es el teoréma de las funciones Lipschitz continuas. Antes de enunciarlo es preciso que definamos Lipschitz continuo.

Decimos que una función $g\colon (S,d) \to (S_1,d_1)$ es Lipschitz continua si existe una constante l, tal que

$$d_1(g(x),g(y)) \leq ld(x,y)\,.$$

Al ínfimo de estas constantes lo llamaré constante de Lipschitz y lo denotaré L.

Denotaré $\rho(X,Y)$ a la distancia en el sentido de Prohorov de las medidas inducidas por las respectivas variables aleatorias X e Y.

Ahora vamos a enunciar el teorema de las funciones Lipschitz continuas

Teorema 1.26. *Supongamos que $g\colon (S,d) \to (S_1,d_1)$ es una función Lipschitz continua en un subconjunto B de S. Entonces*

$$\rho(g(X),g(Y)) \leq \text{máx}(L,1)\rho(X,Y)$$

para toda pareja de variables aleatorias X y Y de (S,d) tales que $P(Y \in B) = 1$.

La prueba de esto conciste en utilizar la definición de la cota L y de la métrica de Prohorov.

1.5 Espacio generalizado de Prohorov

Notemos que toda medida finita es susceptible a ser comparada en el sentido de Prohorov: Sea Q una medida finita, $\frac{Q}{Q(S)}$ es una medida de probabilidad sobre S y por lo tanto está en el espacio de Prohorov sobre S, podemos entonces metrizar la convergencia débil de cualquier medida finita pensando en su distancia en el espacio de las probabilidades y la diferencia entre la medida que le asocian a todo el espacio S, llamaremos la métrica generalizada de Prohorov para medidas finitas sobre S a:

$$\varrho(Q,P) = \text{máx}(\rho(\frac{Q}{Q(S)},\frac{P}{P(S)}),|Q(S)-P(S)|)\,.$$

Es trivial ver qué es métrica, pues el máximo de dos métricas también lo es.

1.5.1 Convergencia de las medidas de Levy

Diremos que una medida Π es medida de Levy si $\int min(1, x^2)\Pi < \infty$

Trabajaremos en las medidas de Lévy. Nosotros ya conocemosal espacio de Prohorov generalizado, además tenemos la siguiente relación 1 a 1 entre medidas finitas que no asignan masa al cero y las medidas de Lévy:

Sea ζ una medida de Lévy, existe una medida finita v tal que $v(\{0\}) = 0$ y

$$\frac{d\zeta}{dv} = 6(1 - \frac{siny}{y}),$$

e inversamente

$$\frac{dv}{d\zeta} = \frac{1}{6(1 - \frac{siny}{y})}$$[3]

además sabemos que una sucesión de medidas de Lévy A_n converge débilmente a A si

1. A_n converge débilmente restringida a todos los conjuntos de la forma $(-\infty, -\epsilon) \bigcup (\epsilon, \infty)$,
2. $\int_{R/\{0\}} x^2 A_n \to \int_{R/\{0\}} x^2 A$.[4]

Definimos la métrica ϱ. Sean A y B medidas de Levy , entonces

$$\varrho(A, B) = \rho(A', B');$$

donde $A' = \frac{1}{6(1 - \frac{siny}{y})} A$ y $B' = \frac{1}{6(1 - \frac{siny}{y})} B$.

Sabemos que esta es en efecto una métrica, pues esta en términos de la métrica generalizada de Prohorov. Falta ver que convergencia en ϱ implica convergencia débil, para probar esto, probaremos algo un poco más fuerte.

Teorema 1.27. *La función de las medidas de Levy a las finitas sin masa en el cero dada por*

$$f(A) = A' \text{ con } A \text{ y } A' \text{ definidas como anteriormente}$$

es una biyección bicontinua en el sentido de la convergencia débil.

[3] véase Fristedt p. 296. [4]

[4] Fuente: comunicación personal con el doctor Victor Rivero

Demostración. Sabemos que es una relación uno a uno e invertible[5] falta mostrar que es continua y tambien su inversa (en terminos de convergencia débil). Para mostrar la continuidad tomemos una sucesión que converga en las medidas de Levy

$$A_n \to A,$$

notemos que la función $g(y) = 6(1 - \frac{siny}{y})$ es contiua en R, que vale cero en cero y es acotada. Sea h una función continua y acotada, claramenta gh es continua y acotada, entonces

$$\int_R hA'_n = \int_{R/\{0\}} hA'_n = \int_{R/\{0\}} hgA_n \to \int_{R/\{0\}} hgA = \int_{R/\{0\}} hA' = \int_R hA',$$

es decir

$$\int_R hA'_n \to \int_R hA'.$$

Inversamente, tomemos una sucesión de medidas finitas convergente A'_n, A', veamos si converge bajo f^{-1}. Tenemos que $\frac{1}{g(y)}$ tiende a infinito cerca del cero y ahí no es continua, notemos también que es de orden $\frac{1}{x^2}$, y por último que si la restringimos a $(-\infty, -\epsilon) \bigcup (\epsilon, \infty)$ entonces sí es continua y acotada. Esto nos permite asegurar la condición 1 para convergencia de medidas de Levy de manera análoga a lo que acabamos de hacer. Para garantizar que se cumple la condicion 2 basta notar que $\frac{x^2}{g(x)}$ es de orden 1, entonces es integrable respecto a nuestra medida finita. Despues de notar esto se procede de manera análoga. Esto termina la prueba. □

Este es un ejemplo de como se puede generalizar la metrica de Prohorov a medidas no necesariamente finitas.

1.6 μ^m

Sea $C = \{x_1..., x_m\}$ un conjunto finito de puntos, con m elementos en un espacio metrico S completo y separable. Definimos al espacio metrico

$$\mu_C^m = \{p \in W : p(C^c) = 0\}$$

Donde W es el conjunto de las medidas finitas sobre S y la metrica es ρ. En esta sección probaremos que

$$\mu_C^m = \mathbb{R}^m$$

[5]véase Fristedt p. 296

donde la igualdad es en el sentido topologico, es decir, son homeomorfos. Para demostrar esto introducimos la funcion

$$f : \mu_C^m \to \mathbb{R}^m$$

tal que si $p \in \mu^m$

$$f(p) = (p(x_1), p(x_2)..., p(x_m)) \in \mathbb{R}^m$$

y si $a = (a_1, ...a_m) \in \mathbb{R}^m$

$$f^{-1}(a) = \sum_{i=1}^{m} a_i \delta_{x_i} \in \mu_C^m$$

Teorema 1.28. *μ_C^m y $\mathbb{R}^m$ son homeomorfos.*

Demostración. Claramente f es una relación 1 a 1. Probaremos que f es bi-continua, probando que manda suceciones convergentes en suceciones convergentes. Sea

$$\delta = min\{d(x, y) : x, y \in C\}$$

- Sea $p_n, p \in \mu_C^m$ tal que $p_n \to p$. Queremos ver que $p_n(x_j) \to p(x_j)$ para toda $j \in \{1, 2...m\}$. Supongamos que para alguna j, $|p_n(x_j) - p(x_j)| > k$ para todo n, para alguna $k > 0$. Claramente esto implica que $\rho(p, p_n) \geq min(k, \delta)$ lo que contradice que $p_n \to p$. La contradiccion viene de suponer que para alguna j $|p_n(x_j) - p(x_j)| > k$ para todo n, para alguna $k > 0$, por lo tanto, para toda j $|p_n(x_j) - p(x_j)| \to 0$ lo que implica que $f(p_n) \to f(p)$.

- Sea $a = (a^1, ...a^m), a_n = (a_n^1..., a_n^m) \in \mathbb{R}^m$. Esto implica que $a_n^j \to a^j$ para toda $j \in \{1, ..., m\}$. Sea $\delta > \epsilon > 0$. Sea N tal que para toda $n > N$, $|a_n^j - a^j| < \frac{\epsilon}{m}$, para toda $j \in \{1, ..., m\}$. Sea $I_n = \{j \in \{1, ..., m\} : a_n^j - a^j > 0\}$. Es claro que para toda $n > N$, el conjunto $\bigcup_{j \in I_n} x_j$ es $\epsilon - conjunto$ de maxima diferencia. Entonces

  $$\rho(f^{-1}(a), f^{-1}(a_n)) \leq f^{-1}(a_n)(\{\bigcup_{j \in I_n} x_j\}) - f^{-1}(a)(\{\bigcup_{j \in I_n} x_j\}^\epsilon)$$

  Pero como $\delta > \epsilon$

  $$f^{-1}(a_n)(\{\bigcup_{j \in I_n} x_j\}) - f^{-1}(a)(\{\bigcup_{j \in I_n} x_j\}^\epsilon) = f^{-1}(a_n)(\{\bigcup_{j \in I_n} x_j\}) - f^{-1}(a)(\{\bigcup_{j \in I_n} x_j\})$$

Pero

$$\sum_{j\in I_n} |a_n^j - a^j| \leq \epsilon$$

por lo tanto

$$\rho(f^{-1}(a), f^{-1}(a_n)) \leq \epsilon$$

Lo que prueba la convergencia.

□

Corolario 1.29. *Las medidas de probabilidad sobre C son homeomorfas al hiperplano* $\{a = (a_1, ...a_m) \in \mathbb{R}^m : \sum_{j=1}^m a_j = 1\}$

Demostración. Basta notar que la imagen inversa del hiperplano bajo f son las medidas de probabilidad. □

1.7 Espacio producto

En esta sección abordaremos superficialmente el tema del espacio producto, no presentaremos demostraciones pero estas son en general sencillas [6]. Trabajaremos en espacios métricos arbitrarios.

Podemos preguntarnos sobre cómo metrizar la convergencia conjunta débil de medidas de probabilidad, es decir, en lugar de preguntarnos por la sucesión X_n en (S_1, m_1) y Y_n en (S_2, m_2), nos preguntamos por (X_n, Y_n) en $(S_1 \times S_2, m)$.

La métrica m que usamos para S puede definirse de varias maneras, la que nosotros usaremos por ser la más maniobrable es

$$m((x_1, y_1)(x_2, y_2)) = \text{máx}(m_1(x_1, x_2), m_2(y_1, y_2))\,.$$

Sea $S = S_1 \times S_2$. Recordemos algunas definiciones básicas, si P es una medida de probabilidad en (S, m), definimos las medidas de probabilidad marginales, P_i en (S_i, m_i) para i=1,2 como

$$P_1(A) = P(A \times S_2), \quad A \in B(S_1)$$
$$P_2(A) = P(A \times S_1), \quad A \in B(S_2)\,.$$

Para fines de definir la topología del producto,

$$G = G_1 \times G_2 = \{(x, y) \colon x \in G_1, y \in G_2\}\,,$$

[6] y se pueden encontrar en otras publicaciones como el libro de Whitt

es abierto si G_i es abierto en S_i con $i = 1, 2$.

En este espacio S, tenemos propiedades interesantes, la primera que observaremos es sobre la separabilidad.

Teorema 1.30. *El espacio S es separable si y sólo si los espacios S_1 y S_2 son separables.*

La prueba es muy sencilla, si S es separable, la proyección del denso numerable en S_i, es denso y numerable. Mientras que si S_1 y S_2 son separables, existe un denso numerable D_i en S_i, definimos el conjunto $D = \{(x, y) : x \in D_1, y \in D_2\}$, este conjunto claramente es denso y numerable.

El siguiente teorema se presentará sin prueba, nos habla de la σ-algebra del espacio producto S.

Teorema 1.31. *Los Borelianos del espacio producto S, $B(S)$, con la topología del producto, es la σ-algebra producto de los Borelianos de los espacios componentes, $B(S) = B(S_1) \times B(S_2)$, si sólo si (S_i, m_i) es separable para $i = 1, 2$.*

1.8 Conexidad

Probaremos que el espacio de las medidas finitas sobre (S,d) completo y separable es conexo.

Teorema 1.32. *El espacio de las medidas finitas es conexo.*

Demostración. Lo que haremos primero es probar que el subconjunto de las medidas finitas con soporte finito, i.e.

$$L = \{m \in W : m(\{x_1, ...x_l\}^c) = 0\}$$

, para algún conjunto (no fijo) finito de puntos $\{x_1, ...x_l \in S\}$ es conexo.

Sea $p \in L$ tal que $p(\{x_1, ...x_l\}^c) = 0$. Queremos probar que la componente conexa de p es L. Sea $q \in L$ tal que $q(\{y_1, ...y_h\}^c) = 0$. Sin perdida de generalidad diremos que hay r elementos en $\{x_1, ...x_l\} \bigcup \{y_1, ...y_h\}$ Notemos que

$$\mu^r_{\{x_1,...x_l\} \bigcup \{y_1,...y_h\}} = \{m \in W : m((\{x_1, ...x_l\} \bigcup \{y_1, ...y_h\})^c) = 1\}$$

es homeomorfo a $\mathbb{R}^r$ y por lo tanto es conexo. Como p y q son elementos de $\mu^r_{\{x_1,...x_l\} \bigcup \{y_1,...y_h\}}$, q esta en la componente conexa de p. Como q era arbitrario, L es la componente conexa de p. Por lo tanto L es conexo.

La segunda parte de la prueba consiste en notar que L es denso en W. Sea m una medida finita sobre S. Sabemos que existe un compacto K tal que $m(k) > 1-\epsilon$. Por lo aprendido en la sección de aproximaciones sabemos que hay una medida finita, p, con soporte finito contenido en K, tal que $\rho(m,p) < \epsilon$. Lo que prueba que L es denso.

Como L es conexo y denso en W, W es conexo. [7]

□

El siguiente teorema es un corolario de lo que acabamos de probar, sin embargo resulta interesante desde el punto de vista de la topología.

Teorema 1.33. *(Teorema del puente) Para todo especio métrico (S,d) completo y separable,*

existe un espacio métrico (E,h) completo, separable y conexo tal que

- $S \subset E$
- $(S,d) = (S,h)$

Demostración. Basta escribir a $x \in S$ como δ_x la medida de Dirac con soporte en x y a E como el conjunto de las medidas con soporte en (S,d) y la métrica ρ. □

1.9 La convergencia débil y la métrica de Prohorov

Definición 1.34. Diremos que una sucesión $\{P_n\}$ de medidas de probabilidad sobre S, converge débilmente a P, si para toda función continua y acotada $f : S \to \mathbb{R}$, se cumple:

$$\lim_{n\to\infty} \int f dP_n = \int f dP\,.$$

Tenemos los siguientes teoremas que asocian el concepto de convergencia débil, con convergencia en el sentido de la métrica de Prohorov.

Teorema 1.35. *Son equivalentes las siguientes afirmaciones:*

1. $P_n \to P$ *débilmente*

[7] Para la prueba de que todo conjunto que contiena a un denso conexo es conexo consulte cualquier libro de topología general, por ejemplo [11]

2. $\lim \int f P_n = \int f dP$ *para todas las funciones acotadas y uniformemente continuas*

3. $\lim sup P_n(F) \leq P(F)$ *para todos los conjuntos cerrados* $F \subset S$

4. $\lim inf P_n(G) \geq P(G)$ *para todos los abiertos* $G \subset S$

5. $\lim P_n(A) = P(A)$ *para todo conjunto* $P-$*continuo* $A \subset S$.

No demostraremos este teorema pero daremos la idea de cada implicación. Para probar que 1 implica 2 se debe notar que las funciones uniformemente continuas son subconjunto de las funciones continuas. 2 implica 3 se prueba usando las funciones $f_\epsilon(x) = \max\{(1 - \frac{d(x,F)}{\epsilon}), 0\}$ que tienden a la indicadora de F cuando $\epsilon \to 0$ y son continuas. 3 implica 4 por complementación. 4 implica 5 porque el interior de un conjunto es abierto y la frontera tiene probabilidad cero en los conjuntos P-continuos. Por último 5 implica 1 pues los conjuntos de la forma $\{f \geq t\}$ son P-continuos.[8]

Ahora estos teoremas muestran la importancia de la métrica de Prohorov.

Teorema 1.36. *Sea* (S, d) *un espacio arbitrario y* $\{P_n\}$ *una sucesión de medidas de probabilidad sobre* S. *Entonces* $\lim_{n\to\infty} \rho(P_n, P) = 0$ *implica que* $P_n \to P$ *débilmente*

Demostración. Definimos $\epsilon_n = \rho(P_n, P) + \frac{1}{n}$. Sea $f > 0$ continua y acotada, como $\epsilon_n > \rho(P_n, P)$ tenemos las siguientes desigualdades:

$$E(f(X_n)) = \int f dP_n = \int_0^{||f||} P_n\{f \geq t\} dt \leq \int_0^{||f||} (P(\{f \geq t\}^{\epsilon_n}) + \epsilon_n) dt$$
$$= \int_0^{||f||} (P(\{f \geq t\}^{\epsilon_n})) dt + \epsilon_n ||f|| \,.$$

Donde X_n induce la medida P_n y X la medida P. La segunda desigualdad es por definición de la métrica de Prohorov y por el hecho de que $\epsilon_n > \rho(P_n, P)$. La anterior desigualdad es para toda n, entonces podemos tomar límites para tener

$$\limsup_{n\to\infty} \int f dP_n \leq lim_{n\to\infty} \int_0^{||f||} P(\{f \geq t\}^{\epsilon_n}\}) dt + \epsilon_n ||f||$$
$$= \int_0^{||f||} P\{f \geq t\} dt = \int f dP \,.$$

[8]Para la demostración completa del teorema: Ethier y Kurtz

De donde para todas las funciones positivas, continuas y acotadas se cumple que:

$$\limsup_{n\to\infty} \int f dP_n \leq \int f dP\,.$$

Sea g continua y acotada, es claro que $||g|| + g$ y $||g|| - g$ son positivas continuas y acotadas, entonces:

$$\limsup_{n\to\infty} \int (||g|| + g) dP_n \leq \int (||g|| + g) dP$$
$$\limsup_{n\to\infty} \int (||g|| - g dP_n) \leq \int (||g|| - g) dP\,.$$

Como P_n y P son medidas de probabilidad podemos despejar de dichas desigualdades

$$\limsup_{n\to\infty} \int g dPn \leq \int g dP$$
$$\limsup_{n\to\infty} \int -g dPn \leq \int -g dP;$$

lo que implica, multiplicando la segunda desigualdad por -1, que

$$\liminf_{n\to\infty} \int g dPn \geq \int g dP\,.$$

Lo que termina la prueba. □

Por último el siguiente teorema nos da, en espacios separables, un si sólo si del teorema que vincula la métrica de Prohorov y la convergencia débil.

Teorema 1.37. *Si S es separable y $P_n \to P$ débilmente entonces* $\lim_{n\to\infty} \rho(P_n, P) = 0$

Demostración. Sea $\epsilon > 0$. Sea $E_1, E_2, \ldots$ una partición con elementos medibles de S, tal que diametro$(E_i) < \frac{\epsilon}{2}$ para toda i. Sea N el menor entero positivo tal que $P(\bigcup_{i=1}^{N} E_i) > 1 - \frac{\epsilon}{2}$ (notemos que $\sum_{i=1}^{\infty} P(E_i) = 1$, la convergencia de dicha serie nos asegura que N existe). Sea g la colección finita de abiertos de la forma $(\bigcup_{i\in I} E_i)^{\frac{\epsilon}{2}}$ donde $I \subset \{1, 2, \ldots, N\}$. Como $P_n \to P$ débilmente, por el teorema anterior tenemos que para todo abierto G de S

$$\liminf P_n(G) \geq P(G)$$

esto nos asegura que existe un n_0 tal que $P(G) \leq P_n(G) + \frac{\epsilon}{2}$ para todo $n > n_0$ y para todo $G \in g$ puesto que g es finito. Sea F cerrado, definimos

F_0 de la siguiente manera:

$$F_0 = \bigcup_{E_i \cap F \neq \emptyset} E_i \,.$$

Se cumple que $F_0^{\frac{\epsilon}{2}}$ está en g, además como la partición tiene diámetro menor que $\frac{\epsilon}{2}$, tenemos que $F_0 \subset F^\epsilon$, entonces

$$P(F) \leq P(F_0^{\frac{\epsilon}{2}}) + \frac{\epsilon}{2} \leq P_n(F_0^{\frac{\epsilon}{2}}) + \epsilon \leq P_n(F^\epsilon) + \epsilon$$

de donde para todo $\epsilon > 0$ se cumple que $\rho(P, P_n) < \epsilon$ para todo $n \geq n_0$, lo que implica convergencia en el sentido de Prohorov. $\square$

Ahora hablemos de cómo podemos generar, a partir de funciones continuas, nuevas sucesiones convergentes, conociendo una sucesion que converge.

El siguiente teorema es conocido como el teorema simple de la funcion continua y es muy intuitivo.

Teorema 1.38. *Si $X_n \to X$ débilmente en (S, d) y $g\colon (S, d) \to (S_1, d_1)$ es continua, entonces*

$$g(X_n) \to g(X) \ \textit{débilmente en} \ \ (S_1, d_1)\,.$$

Para probarlo veremos que:

$$E[f(g(X_n))] \to E[f(g(X))]\,.$$

Para toda f, función continua y acotada sobre (S_1, d_1). Lo anterior sucede pues para cualquier función f continua, sobre (S_1, d_1) la composición $f \circ g$ es una función continua y acotada sobre (S, d), lo que prueba el resultado.

Ahora enunciaremos 2 maneras de generalizar el teorema simple de la función continua

Teorema 1.39. *Si $X_n \to X$ débilmente y $g\colon (S, d) \to (S_1, d_1)$ es medible con $P(X \in \mathrm{disc}(g)) = 0$ entonces*

$$g(X_n) \to g(X) \ \textit{débilmente en} \ (S_1, d_1).$$

Lo que nos dice el teorema anterior, es que si la funcion g no es continua, pero es improbable que la variable aleatoria X caiga en un punto de discontinuidad, entonces se preserva la convergencia.

El siguiente teorema incluye como caso particular al anterior, sin embargo entender lo que nos dice el que acabo de plantear, nos ayuda a entender el siguiente.

Teorema 1.40. *Sea g y g_n funciones medibles $(S, d) \to (S_1, d_1)$, donde (S_1, d_1) es separable. Llamemos R al conjunto de puntos x en S, tales que $g_n(x_n)$ no converge a $g(x)$, para alguna sucesión x_n que converge a x. Si $X_n \to X$ débilmente y $P(X \in R) = 0$ entonces*

$$g_n(X_n) \to g(X) \ \textit{débilmente en} \ \ (S_1, d_1)\,.$$

Notemos que si $g_n = g$, $R = \mathrm{disc}(g)$ y tenemos el teorema anterior.

Para conocer mas de las aplicaciones y las pruebas de los teoremas se puede consultar la referencia bibliografica [9].

1.10 Sumario de las propiedades topologicas de las medidas finitas

El espacio de las medidas finitas sobre los reales con la topologia inducida por la convergencia débil cumple:

1. Es metrizable
2. Es conexo
3. Es separable
4. Es completo
5. Es primero y segundo numerable
6. Es Housdorff
7. No es localmente compacto

Capítulo 2

Espacio de Skorohod

(Ver [9]) Las funciones Cadlag, (continuas por la derecha con límite por la izquierda) son de gran interés, ya que se usan para modelar los procesos con trayectorias discontinuas, el espacio de las funciones Cadlag con dominio en el intervalo $[0, T]$, al que llamaremos $D[0, T]$ es conocido como el espacio de Skorohod.

Darle una métrica a este espacio y explorar sus consecuencias es a lo que dedicaré este capítulo. En este orden de ideas la primera función que se me ocurre para metrizar este espacio es la métrica del supremo, pues sabemos que funciona muy bien con el espacio de las funciones continuas. Sin embargo para este espacio no nos será de gran utilidad. Para convencernos de ello basta estudiar algunos ejemplos.

Primero recordemos la métrica del supremo, sea $x \in D([0, T], \mathbb{R})$ entonces

$$||x - y|| = \sup_{0 \leq t \leq T} |x(t) - y(t)| .$$

Como ya mencioné esta métrica en muy útil cuando las funciones son continuas, sin embargo en el espacio $D([0, T], \mathbb{R})$ esta métrica no puede determinar correctamente la cercanía de algunas funciones, un ejemplo muy sencillo es el siguiente:

Ejemplo 2.1. Sea $X = I_{[0,1/2)}$ y sea $X_n = I_{[0,1/2+1/n)}$. Claramente cuando n tiende a infinito $\frac{1}{2} + \frac{1}{n}$ tiende a $\frac{1}{2}$ por lo que nos gustaría que X_n convergiera a X, sin embargo $||X - X_n|| = 1$ para todo n, por lo tanto según la métrica del supremo esta sucesión no converge a X.

Probablemente este era el tipo de ejemplos que tenia Skorohod en mente cuando diseño las métricas que estudiaremos a continuación. Estas son mucho más adecuadas para analizar este espacio, pues se portan bien incluso en ejemplos como el anterior.

2.1 La métrica $J1$

En esta sección definiré la métrica $J1$ sobre el espacio D, (espacio de las funciones Cadlag), para funciones cuyo dominio sea un intervalo finito $[0,T]$ y codominio $\mathbb{R}$, denotaremos dicho espacio $D([0,T],\mathbb{R}^k)$.

Definimos la métrica J_1 en intervalos $[0,T]$ como sigue: sean $x, y \in D([0,T],\mathbb{R})$, sea Λ el conjunto de homeomorfismos estrictamente crecientes del $[0,T]$ en sí mismo, entonces

$$d_{j_1T}(x,y) = \inf_{\lambda\in\Lambda} \max\{||x\circ\lambda - y||, ||e-\lambda||\} .$$

Teorema 2.2. *Es una métrica en $D[0,T]$.*

Demostración.

1. d_J es no negativa.

2. $d_{j1}(x,y) = 0$ si y sólo si $x = y$.

   Como $e \in \Lambda$ es claro que si $x = y$

   $$d_{j1}(x,y) = 0,$$

   ahora, si $d_{j1}(x,y) = 0$ tenemos que el máximo de $||x\circ\lambda - y||$ y $||e-\lambda||$ es cero, por lo tanto $\lambda(t) = e(t)$ para todo t, lo cual implica que $||x\circ\lambda - y||| = ||x - y|| = 0$ de donde $x = y$.

3. $d_{j1}(x,y) = d_{j1}(y,x)$.

   Notemos que el inverso de todas las λ, que denotaremos λ^{-1}, tambien está en Λ, además $||e-\lambda|| = ||e-\lambda^{-1}||$ y $||x\circ\lambda - y|| = ||y\circ\lambda^{-1} - x||$ por lo tanto el conjunto de posibles valores con $\lambda \in \Lambda$ de $\max\{||x\circ\lambda - y||, ||e-\lambda||\}$ es el mismo de $\max\{||y\circ\lambda - x||, ||e-\lambda||\}$ por lo tanto $d_{j1}\ (x,y) = d_{j1}(y,x)$.

4. Desigualdad del triángulo.

   Notemos que toda λ en Λ se puede ver como la composición de dos elementos de Λ, pues en particular la identidad es elemento deΛ, es decir $\lambda_1 \in \Lambda$ si solo si existe $\lambda_2, \lambda_3 \in \Lambda$ tal que $\lambda_1 = \lambda_2 \circ \lambda_3$.

   Para esta prueba tambien usaremos que la métrica del supremo tiene la propiedad subaditiva, por ser métrica, y finalmente que el ínfimo con $\lambda_1, \lambda_2 \in \lambda$ es menor que si fijamos $\lambda_2 = e$ y hacemos el ínfimo sobre λ_1.

Al usar todas las observaciones anteriores tenemos que para $x, y, z \in D$

$$\begin{aligned}
d_{j_1T}(x,y) &= \inf_{\lambda\in\Lambda} \max\{||x\circ\lambda - y||, ||e-\lambda||\} \\
&= \inf_{\lambda_1,\lambda_2\in\Lambda} \max\{||x\circ\lambda_1\circ\lambda_2 - y||, ||e-\lambda_1\circ\lambda_2|| \\
&\leq \inf_{\lambda_1,\lambda_2\in\Lambda} \max\{||x\circ\lambda_1\circ\lambda_2 - z\circ\lambda_2 + z\circ\lambda_2 - y||, ||e-\lambda_2+\lambda_2-\lambda_1\circ\lambda_2||\} \\
&\leq \inf_{\lambda_1,\lambda_2\in\Lambda} \max\{||x\circ\lambda_1\circ\lambda_2 - z\circ\lambda_2||, ||\lambda_2-\lambda_1\circ\lambda_2||\} \\
&\qquad + \inf_{\lambda_2\in\Lambda} \max\{||z\circ\lambda_2 - y||, ||e-\lambda_2||]\}
\end{aligned}$$

Finalmente tomamos $\lambda_2 = e$ en el primer sumando, entonces

$$\begin{aligned}
&\leq \inf_{\lambda_1\in\Lambda,\lambda_2=e} \max\{||x\circ\lambda_1\circ e - z\circ e||, ||e-\lambda_1\circ e||\} \\
&\qquad + \inf_{\lambda_2\in\Lambda} \max\{||z\circ\lambda_2 - y||, ||e-\lambda_2||\} \\
&= \inf_{\lambda_1,\in\Lambda} \max\{||x\circ\lambda_1 - z||, ||e-\lambda_1||\} \\
&\qquad + \inf_{\lambda_2\in\Lambda} \max\{||z\circ\lambda_2 - y||, ||e-\lambda_2||\} \\
&= d_J(x,z) + d_J(z,y)\,,
\end{aligned}$$

y por lo tanto

$$d_J(x,y) \leq d_J(x,z) + d_J(z,y)\,.$$

□

Notemos que $||x-y|| \geq d_{j_1}(x,y)$, para probarlo basta tomar $\lambda = e$.

Un dato interesante es que esta métrica restringida a las funciones continuas sobre un intervalo acotado es equivalente a la métrica del supremo si las funciones que estamos comparando son continuas.

Demostración. sean x, x_n continuas, si $|x - x_n| \to 0$, entonces

$$d_{j_1T}(x,x_n) = \inf_{\lambda\in\Lambda} \max\{||x\circ\lambda - x_n||, ||e-\lambda||\} \to 0\,,$$

basta tomar $\lambda = e$.

Inversamente si $d_{j_1T}(x,x_n) \to 0$, para toda $\epsilon > o$ existe una N tal que existe un homeomorfismo λ tal que para toda $n > N$ $|x\circ\lambda - x_n| < \epsilon$ y $|\lambda - e| < \epsilon$, pero por continuidad de x $|x\circ\lambda - x| < \epsilon$.

$$|x - x_n| \leq |x - x\circ\lambda| + |x\circ\lambda - x_n|$$

lo que termina la prueba. □

Para mostrar la utilidad de esta métrica volveremos a analizar la sucesión del ejemplo 1, en el siguiente ejemplo:

Ejemplo 2.3. La sucesión $X_n = I_{[0,1/2+1/n)}$ converge a $X = I_{[0,1/2)}$ en $D[0,1]$ con la metrica d_{j_1}.

(Se recomienda ampliamente que el lector haga un dibujo) Definimos $\lambda_n = \frac{nx}{n+2} I_{[0,\frac{1}{2}+\frac{1}{n})} + [\frac{nx}{n-2} + \frac{2}{n-2}] I_{[\frac{1}{2}+\frac{1}{n},1]}$ que es trazar una recta del $(0,0)$ al $(\frac{1}{2}+\frac{1}{n},\frac{1}{2})$ y pegarla con la recta que va del $(\frac{1}{2}+\frac{1}{n},\frac{1}{2})$ al $(1,1)$, claramente $I_{(0,\frac{1}{2}+\frac{1}{n})=I_{(0,\frac{1}{2})}} \circ \lambda_n$ y también es muy claro que conforme n tiende a infinito, las pendientes del las rectas que componen λ_n tienden a uno, por lo que λ_n tiende a la identidad. En conclusión, $d_{j_1 1}(X, X_n) = \inf_{\lambda\in\Lambda} \max\{||X \circ \lambda - X_n||, ||e - \lambda||\} \leq ||e - \lambda_n|| = \frac{1}{n}$ por lo que $X_n \to X$ en $D[0,1]$. De donde en efecto esta métrica nos explica mejor lo que pasa en este espacio, pues toma en cuenta las discontinuidades en las trayectorias.

Al menos en este ejemplo la métrica J_1 es más adecuada que la del supremo. El éxito de nuestra métrica sobre la del supremo se debe a que pedimos una convergencia en cada t distinta a la habitual, como el siguiente teorema nos muestra. Definimos $x(t^-)$ como $\lim_{n\to\infty} x(t_n)$ para toda sucesión $t_n < t$, si $x(t^-)$ existe, es decir es único, decimos que x tiene límite por la izquierda en t. Las funciones Cadlag tienen límite por la izquierda en todos los puntos.

Teorema 2.4. *Si $d_{j1}(x_n, x) \to 0$ en $D[0,T]$ entonces para todo t tal que $0 < t < T$ existe una sucesión t_n tal que $t_n \to t$, $x_n(t_n) \to x(t)$, $x_n(t_n-) \to x(t-)$ y $x_n(t_n) - x_n(t_n-) \to x(t) - x(t-)$.*

Demostración. Como $d_J(x_n, x) \to 0$ existe una sucesión de funciones λ_n tal que $|x \circ \lambda_n - y|| \to 0$ y $||e - \lambda_n|| \to 0$, esto es facil de ver, supongamos que $d_{j1}(X_n, X) = k_n$, por ser un ínfimo sabemos que existe una λ_n tal que

$$\max\{||x_n \circ \lambda_n - x||, ||\lambda_n - e||\} < \frac{1}{n} + k_n \,.$$

Esta sucesión, $\{\lambda_n\}_{n\geq 1}$, cumple lo que queríamos. Sea $t_n = \lambda_n(t)$ como $\lambda_n \to e$ entonces $t_n \to t$ y además como $||x_n \circ \lambda_n - x|| \to 0$ tenemos que en efecto $x_n(t_n) \to X(t)$, y finalmente sea S^m una sucesión que crece a t, entonces $\lambda_n(S^m) \uparrow \lambda_n(t) = t_n$ por que λ_n es creciente y continua. Entonces $x_n(\lambda_n(S^m)) \to x_n(t_n^-)$. Por lo tanto

$$\begin{aligned} |x_n(t_n^-) - x(t^-)| \leq |x_n(t_n^-) - x_n(\lambda_n(S^m))| + |x_n(\lambda_n(S^m)) - x(S^m)| \\ + |x(S^m) - x(t^-)| \end{aligned}$$

donde, por construcción los tres terminos tienden a cero, lo que nos permite concluir que $|x_n(t_n^-) - x(t^-)| \to 0$. De estas dos se concluye la tercera aseveración del teorema $x_n(t_n) - x_n(t_n-) \to x(t) - x(t-)$. $\square$

Lo que acabamos de aprender es la base del funcionamiento de la métrica J1, esta función nos dice cuando dos trayectorias se parecen, tomando en cuenta el comportamiento de sus saltos, nos asegura que una sucesión de funciones converge a una función g solo si los saltos de la sucesión convergen a los de la función límite, el siguiente ejemplo ilustra esa idea.

Ejemplo 2.5. Sea $k < T, k \in naturales$ definimos $X_n = \sum_1^k iI_{[i-1+\frac{1}{n}, i+\frac{1}{n})}$ y $X = \sum_1^k iI_{[i-1,i)}$, en este ejemplo todas las funciones tienen k saltos, claramente el $k-$ésimo salto de cada sucesión, converge al $k-$esimo salto, pues la sucesión $\frac{1}{n} \to 0$.

Usando la una técnica similar a la del ejemplo 2, definimos la familia de funciones λ_n de la siguiente forma, λ_n es igual a la función cuya gráfica es el segmento de línea recta que unen el $(0,0)$ con $(1+\frac{1}{n}, 1)$ pegado con el que va de $(1+\frac{1}{n}, 1)$ a $(k+\frac{1}{n}, k)$ pegado con el que une $(k+\frac{1}{n}, k)$ con (T,T).

Entonces la distancia según J1 entre X_n y X es menor que $\frac{1}{n}$ pues $||\lambda_n - e|| = \frac{1}{n}$, lo que muestra explícitamente la convergencia.

Veamos ahora qué pasa si los saltos no convergen.

Ejemplo 2.6. Sea $X_n = \text{máx}\{0, 1 - nd(x,1)\}$ y sea $X = I_{\{(1)\}}$ intuitivamente $X_n \to X$, así que nos gustaría que se diera la convergencia en $D[0,T]$. Sin embargo con la métrica J1 mostraremos que no hay convergencia.

Sea $\lambda \in \Lambda$, Tenemos 2 casos:

1. Existe $x_0 \neq 1$, tal que $\lambda(x_0) = 1$, entonces se cumple que $d(X_n \circ \lambda(x_0), X(x_0)) = d(X_n(1), 0) = d(1,0) = 1$.

2. Si no existe Existe $x_0 \neq 1$, tal que $\lambda(x_0) = 1$, como λ es homeomorfismo, se cumple que $\lambda(1) = 1$. Entonces tenemos que $d(X_n \circ \lambda(1^-), X(1^-)) = d(X_n(1^-), 0) = d(1^-, 0) = 1$

como este análisis fue para toda λ, concluimos que $D_{J1}(X, X_n) \geq 1$, de donde no convergen según la métrica J_1.

Sorprendentemente esta métrica no nos dijo lo que queríamos escuchar en ese caso, pero notemos que nosotros definimos nuestra métrica sobre las funciones Cadlag, y claramente $X = I_{\{(1)\}}$ no es continua por la derecha, tambien notamos que era predecible que fallara por el criterio de los saltos, hagamos más ejemplos para ver qué es lo que es lo que está pasando.

Tomemos un ejemplo similar, pero donde todas las funciones sean Cadlag.

Ejemplo 2.7. Sea $X_n = \max(0, 1 - [n][d(x, [0, 1))]$ y sea $X = I_{\{[0,1)\}}$.

Un razonamiento analogo al del ejemplo anterior, nos permite ver que no hay convergencia.

Una vez más no salió lo que nos hubiera gustado, ahora sí estabamos en el caso en el que ambas trayectorias son Cadlag, sin embargo no funcionó. La razón vuelve a ser que los saltos no coinciden, lo que nos dice esto es que esta métrica es demasiado fuerte para este tipo de ejemplos, si queremos que se de la convergencia tendríamos que usar otra métrica, cosa que haremos más adelante.

2.2 La métrica M_1

Tendremos que usar algo nuevo para comparar funciones cuyos saltos no convergan (en particular nos gustaría que la continua y la discontinua del ejemplo que estudiamos converjan), antes de definir esta nueva métrica, tenemos que definir la gráfica completa de una función.

La grafica completa de la función $x \in D$, es el conjunto de puntos

$$\Gamma_x = \{(z, t) \in \Re^2 : z = \alpha x(t^-) + (1 - \alpha)x(t)\,, \quad \text{para algún } \alpha \in [0, 1]\}\,.$$

Lo que se está haciendo es incluir en la gráfica las rectas verticales que unen las discontinuidades, en los puntos de continuidad $x(t^-) = x(t)$ no se agrega ningún punto. La gráfica completa es un subconjunto conexo de R^2, que contiene la gráfica de la función x, y las rectas que unen los puntos de discontinuidad.

Antes de definir la métrica M_1 tenemos que definir el siguiente orden, lo llamaremos orden en la gráfica Γ_x

$$(z_1, t_1) \leq (z_2, t_2) \text{ si } \; t_1 < t_2 \qquad \text{ó}$$

$$t_1 = t_2 \;\text{ y }\; |x(t_1-) - z_1| \leq |x(t_2-) - z_2| = |x(t_1-) - z_2|\,.$$

El orden se puede describir de la siguiente manera, ponemos un hilo sobre la gráfica y lo atamos al principio, es decir a $x(0)$, un punto x es menor que otro y si para llegar caminando sobre el hilo a y, tienes que pasar sobre x.

Ahora, una parametrización de la gráfica completa es una función $p : [0, 1] \subset \mathbb{R} \to \Gamma_x \subset \mathbb{R}^2$, que manda al intervalo $[0, 1]$ en (u, r), donde u es el componente espacial y r el temporal de la gráfica y es creciente en el sentido del orden en la gráfica Γ_x.

Claramente hay más de una parametrización para toda gráfica completa, por ejemplo si nuestra funcion es $x = I_{[\frac{1}{2},1)}$, una posible parametrización es:

$$u_1 = 2tI_{[0,\frac{1}{4})} + \frac{1}{2}I_{[\frac{1}{4},\frac{3}{4})} + (2t-1)I_{[\frac{3}{4},1]}\,, r_1 = (2t - \frac{1}{2})I_{[\frac{1}{4},\frac{3}{4})} + I_{[\frac{3}{4},1)}\,,$$

mientras que otra es:

$$u_2 = tI_{[0,\frac{1}{2})} + \frac{1}{2}I_{[\frac{1}{2},\frac{3}{4})} + (2t-1)I_{[\frac{3}{4},1)}, r_2 = (4t-2)I_{[\frac{1}{2},\frac{3}{4})} + I_{[\frac{3}{4},1)}.$$

Llamemos $\Pi(x)$ al conjunto de todas las parametrizaciones crecientes de $\Gamma(x)$.

Ahora sí estamos en condiciones de definir la métrica M_1

Sean $x, y \in D$

$$d_{M1}(x,y) = \inf_{(u_1,r_1)\in\Pi(x),(u_2,r_2)\in\Pi(y)} \max(|u_1 - u_2|, |r_1 - r_2|)$$

Lema 2.8. *Si x, y son continuas esta métrica es equivalente a la del supremo.*

Demostración. Para ver esto primero suponemos convergencia en terminos del supremo, basta tomar parametrizaciones que tengan parte temporal igual a la identidad. Inversamente tambien es suficiente tomar parametrizaciones que tengan parte temporal igual a la identidad y notar que bajo esas parametrizaciones, la parte espacial es la distancia en terminos del supremo, y que el ínfimo se alcanza cuando las 2 parametrizaciones corren a la misma velocidad en la parte temporal. □

Aun no probaremos que M_1 en efecto es métrica, esto se hará en el apendice 1 (Topologia fuerte y débil), solo adelantaremos algunas cosas que nos permitiran creer que en efecto es métrica.

Claramente es positiva, $d_{M1}(x,x) = 0$, basta tomar la misma parametrización y es reflexiva, es decir $d_{M1}(x,y) = d_{M1}(y,x)$, pues la métrica del supremo lo es, lo único complicado de probar es que cumple la desigualdad del triángulo. Entonces por el momento asumimos que es métrica.

Revisemos el ejercicio que se complicaba con la métrica J_1, para ver si en efecto esta métrica es menos fuerte y permite la convergencia.

Ejemplo 2.9. sea $X_n = \max(0, 1 - [n][d(x, [\frac{1}{2}, 1))]$ y sea $X = I_{\{[\frac{1}{2},1)\}}$ para probar que hay convergencia en el sentido de M_1, vamos a exhibir una parametrización de γ_x y una para la gráfica de cada X_n, la cual es completa, por ser X_n continua, tales que hagan $max(|u_x - u_{x_n}|, |r_x - r_{x_n}|) \to 0$, lo que asegura que $d_{M1}(x_n, x) \to 0$.

Construimos nuestras parametrizaciones de la siguiente manera: sea n fijo, la gráfica de X_n es la constante cero en $[0, \frac{1}{2} - \frac{1}{n}]$, luego la recta que va del punto $(\frac{1}{2} - \frac{1}{n}, 0)$ a $(\frac{1}{2}, 1)$ y finalmente la constante 1 en el intervalo $(\frac{1}{2}, 1]$, mientras que la gráfica completa de x es la constante 0 en el intervalo $(0, \frac{1}{2})$. Luego la recta vertical que va del punto $(\frac{1}{2}, 0)$ a $(\frac{1}{2}, 1)$ y finalmente la constante 1 en el intervalo $(\frac{1}{2}, 1]$. Procederemos de la siguiente manera, la parametrización de la gráfica de X_n manda el intervalo $[0, \frac{1}{3})$ en la recta que une el $(0, 0)$ con el $(0, \frac{1}{2} - \frac{1}{n})$, manda al intervalo $[\frac{1}{3}, \frac{2}{3})$ en la recta que va del punto $(\frac{1}{2} - \frac{1}{n}, 0)$ a $(\frac{1}{2}, 1)$ y el $[\frac{2}{3}, 1]$ en la recta que une el punto $(\frac{1}{2}, 1)$ con el punto $(1, 1)$. Construimos la parametrización de Γ_x mandando el intervalo $[0, \frac{1}{3})$ en la recta que une el $(0, 0)$ con el $(0, \frac{1}{2})$, al intervalo $[\frac{1}{3}, \frac{2}{3})$ en la recta que va del punto $(\frac{1}{2}, 0)$ a $(\frac{1}{2}, 1)$ y el $[\frac{2}{3}, 1]$ en la recta que une el punto $(\frac{1}{2}, 1)$ con el punto $(1, 1)$.

Con estas parametrizaciones notamos que la distancia en la coordeada espacial es cero, pues en los tres segmentos de ambas parametrizaciones se comporta igual, primero es la constante cero, luego crece linealmente en la segunda región de cero a uno, y finalmente es la constante 1 en la última región. en la coordenada temporal notamos que en la primera región, ambas empiezan en el mismo punto pero la segunda crece más rápido hasta que en $\frac{1}{3}$ la primera vale $\frac{1}{2} - \frac{1}{n}$ y la segunda $\frac{1}{2}$, luego la primera crece en la segunda región linealmente, de $\frac{1}{2} - \frac{1}{n}$ a $\frac{1}{2}$, mientras que la segunda permanece constante, y en la tercera región ambas son constantes, por lo que el máximo de las distancias se alcanza en $\frac{1}{3}$ y es $\frac{1}{n}$, por lo que para esta pareja de parametrizaciones tenemos $max(|u_x - u_{x_n}|, |r_x - r_{x_n}) = \frac{1}{n}$.

Como esto fue para n arbitraria obtenemos que $d_{M1}(X_n, X) \to 0$.

Hemos logrado construir una métrica que permite la convergencia de algunas continuas a discontinuas, cosa que no pasaba con J_1, algo que nos podemos preguntar es si convergencia en J_1 implica convergencia en M_1, el siguiente teorema nos asegura eso y algo más.

Teorema 2.10. *Para toda pareja $x_1, x_2 \in D$, se cumple que $d_{J_1}(x_1, x_2) \geq d_{M_1}(x_1, x_2)$.*

Demostración. Nosotros realizaremos la prueba para funciones con una cantidad finita de discontinuidades, pero por densidad de estas el resultado se expande a todas las funciones en D. (Esta afirmación se prueba en la sección 3.3)

Sea $\lambda \in \Lambda$, construiremos una pareja de parametrizaciones en $\Pi_s(x_1)$ y $\Pi_s(x_2)$ respectivamente, tales que $\max |x_1 \circ \lambda - x_2| \, |\lambda - e| = \max(|u_1 - u_2||r_1 - r_2|)$.

Sea t_n la sucesión de los tiempos en los que hay una discontinuidad en la funcion x_1, esta sucesión está ordenada de menor a mayor, i.e. $t_i < t_{i+1}$.

A esta sucesión le asociaremos otra de intervalos cerrados también ordenados, es decir, a t_i le asignamos el intervalo $[a_i, b_i]$, tal que para todo i, $[a_i, b_i] \subset [0, 1]$, $a_{i-1} < b_{i-1} < a_i < b_i < a_{i+1} < b_{i+1}$. Entonces definimos la parte temporal de nuestra parametrización como $r_1(s) = t_n$ para todo $s \in [a_n, b_n]$, y la función $r_1(s)$ en el resto del intervalo $[0, 1]$, la definimos como la interpolacion lineal de lo ya definido. Esto hace a $r_1(s)$ continua y no decreciente.

Ahora, definimos $r_2(s) = \lambda \circ r_1(s)$, y la coordenada espacial la definimos para ambos índices como $u_1(s) = x_1(r_1)$ y $u_2(s) = x_1(r_2)$, para todas las s, que no están en ningún $[a_i, b_i]$, y el resto lo definimos por interpolación lineal.

Claramente esta es una parametrización permitida, además $|x_1 \circ \lambda - x_2| = |u_1 - u_2|$ y $|\lambda - e| = |r_1 - r_2|$ de donde en efecto.

$$\max |x_1 \circ \lambda - x_2| \, |\lambda - e| = \max(|u_1 - u_2||r_1 - r_2|) \, .$$

Como esto fue para toda $\lambda \in \Lambda$, llamemos A al conjunto de todos los valores posibles de $\max |x_1 \circ \lambda - x_2| \, |\lambda - e|$ con $\lambda \in \Lambda$, llamemos B al conjunto de todos los valores posibles de $\max(|u_1 - u_2| \, |r_1 - r_2|)$ con $(u_1, r_1) \in \Pi_s(x_1), (u_2, r_2) \in \Pi_s(x_2)$.

Entonces $A \subset B$, entonces $\inf[A] \geq \inf[B]$, pero $d_{J_1}(x_1, x_2) = \inf[A]$ y $d_{M_1}(x_1, x_2) = \inf[B]$; de donde en efecto

$$d_{J_1}(x_1, x_2) \geq d_{M_1}(x_1, x_2) \, .$$

□

2.3 Regularidad de D

Recordemos que estamos trabajando en el conjunto de las funciones, de un compacto contenido en $\mathbb{R}$ en $\mathbb{R}$, en esta primera parte del capítulo y en $\mathbb{R}^k$ más adelante, continuas por la derecha, con límite por la izquierda. En esta sección observaremos las concecuencias de estas características.

Sea $x \in D$ denotaremos disc(x) al conjunto de tiempos en los que hay una discontinuidad en la funcion x, i.e

$$\mathrm{disc}(x) = \left\{t \in [0,T] : x(t) \neq x(t^-)\right\}$$

además,

$$\mathrm{disc}(x,\epsilon) = \{t \in [0,T] : x(t) \neq x(t^-)\; |x(t) - x(t-)| > \epsilon\}\,.$$

Claramente

$$\mathrm{disc}(x,\epsilon) \subset \mathrm{disc}(x)\,.$$

Teorema 2.11. *Para todo $x \in D$ y $\epsilon > 0$ el conjunto disc(x,ϵ) es finito.*

Demostración. Supongamos que card$\{\mathrm{disc}(x,\epsilon)\}$ es infinito, por ser un conjunto infinito contenido en un compacto tiene al menos un punto de acumulación, supongamos que $t_n \in \mathrm{disc}(x,\epsilon)$ converge por la izquierda a t,el caso inverso es análogo, fijémonos entonces en $x(t^-)$, nosotros sabemos que existe pues x es Cadlag, esto implica que existe una $\delta > 0$ tal que para todo $\tau \in (t-\delta, t)$ se cumple que $|x(\tau) - x(t^-)| < \frac{\epsilon}{2}$, en particular como $t_n \in (t-\delta,t)$ para todos los índices excepto una cantidad finita, tenemos que

$$|x(t_n) - x(t^-)| < \frac{\epsilon}{2}\,,$$

además, definimos τ_{mn} una sucesión tal que $\tau_{mn} \uparrow t_n$, si $t_n \in (t-\delta,t)$ también $\tau_{mn} \in (t-\delta,t)$ para todos los índices excepto un número finito, por lo tanto

$$|x(\lim_{m\to\infty} \tau_{mn}) - x(t)| = |x(t_n^-) - x(t)| < \frac{\epsilon}{2}\,,$$

ahora, recordemos que $|x(t_n) - x(t_n^-)| > \epsilon$, pero por la desigualdad del triángulo tenemos que

$$|x(t_n)-t(t_n^-)| = |x(t_n)-x(t^-)+x(t^-)-t(t_n^-)| \leq |x(t_n)-x(t^-)|+|x(t^-)-t(t_n^-)|\,;$$

pero al menos para una infinidad de índices

$$|x(t_n) - x(t)| + |x(t) - t(t_n^-)| < \epsilon\,,$$

esto implica que $\epsilon > |x(t_n) - x(t_n^-)| > \epsilon$, lo cual claramente una contradicción, que viene de suponer que hay una infinidad de discontinuidades con saltos mayores que ϵ, por lo que

$$\mathrm{card}\{\mathrm{disc}(x,\epsilon)\} \quad \text{es finito.}$$

□

Notemos que el teorema anterior implica que:

Corolario 2.12. disc(x) *es a lo más numerable.*

Demostración. Pues

$$\mathrm{disc}(x,\epsilon) = \bigcup_{n\in N} \mathrm{disc}(x, \frac{1}{n})\,.$$

□

Introduzcamos ahora dos definiciones que nos serán útiles:

Decimos que una función es constante a pedazos o escalonada, si card$\{\mathrm{disc}(x)\} < \infty$ y es constante en $[0,T]$-disc(x).

Definimos el módulo de continuidad de la función x sobre un conjunto $A \subset [0,T]$ como

$$l(x,A) = \sup_{t\in A}(||x(t_1) - x(t_2)||)\,.$$

Ahora podemos enunciar otro teorema de gran interés, que nos permite notar lo bien portadas que son las funciones Cadlag:

Teorema 2.13. *Las funciones escalonadas son densas en D con la métrica del supremo.*

Demostración. Sea $x \in D$ y $\epsilon > 0$, construiremos una funcion escalonada, x_e tal que $||x - x_e|| < \epsilon$, donde $||\cdot||$ es la métrica del supremo.

Consideremos el conjunto disc(x,ϵ), sabemos que es un conjunto finito, igual a $\{0 \le t_0 < t_1 < \ldots < t_k \le T\}$. Definimos $t_-(t_i)$ y $t_-(t_i)$, tales que $t_-(t_i) < t_i < t_+(t_i)$ y

$$l(x,(t_-(t_i),t_i)) < \epsilon \text{ y } \; l(x,[t_i,t_+(t_i))) < \epsilon\,.$$

Esto implica necesariamente que

$$(t_-(t_i),t_i)) \cap \mathrm{disc}(x,\epsilon) = [t_i,t_+(t_i)) \cap \mathrm{disc}(x,\epsilon) = \emptyset\,.$$

Notemos que estos $t_-(t_i)$ y $t_-(t_i)$ existen por ser la función Cadlag. Nuestra función escalonada valdrá $x(t_i-)$ en el intervalo $(t_-(t_i),t_i)$ y $x(t_i)$ en el $[t_i,t_+(t_i))$, en el resto de l intervalo $[0,T]$ la definimos de la siguiente manera:

Notemos que la funcion no está definida en los intervalos cerrados $[t_+(t_i),t_-(t_{i+1})]$, tomando $t_0 = 0$ y $t_{n+1} = T$, cada uno de estos intervalos lo analizamos de la siguiente manera, si

$$l(x,[t_+(t_i),t_-(t_{i+1})]) < \epsilon\,,$$

la función es constante en ese intervalo y vale $x(t)$ para algún t fijo $t \in [t_+(t_i), t_-(t_{i+1})]$.

Supongamos que

$$l(x, [t_+(t_i), t_-(t_{i+1})]) > \epsilon\,.$$

Como $[t_+(t_i), t_-(t_{i+1})] \cap \mathrm{disc}(x, \epsilon) = \emptyset$ para todo $t \in [t_+(t_i), t_-(t_{i+1})]$ existe un abierto (a, b) que contiene a t, tal que

$$l(x, (a, b)) < \epsilon\,.$$

Notemos que la unión de estos abiertos para todo t en el intervalo forman una cubierta abierta y por ser $[t_+(t_i), t_-(t_{i+1})]$ compacto, existe una subcubierta finita, tomamos esta subcubierta finita y por medio de intercecciones hacemos una cubierta disjunta y finita. La función será constante en cada elemento de la cubierta y tomará el valor de $x(t)$ para algún t que pertenezca al elemento de la subcubierta.

Por construcción nuestra función x, cumple que $l(x, A) < \epsilon$ para todo intervalo donde definimos x_s continua, además como en cada uno de esos intervalos $x_s(t) = x(t_0)$ para algún $t_0 \in A$ concluimos que

$$||x - x_s|| < \epsilon\,;$$

el hecho de que x y ϵ eran arbitrarios termina la prueba. □

Corolario 2.14. *Las funciones escalonadas son densas en* (D, M_1) *y* (D, J_1) *.*

Demostración. Sabemos que $||x-x_s|| \geq d_j(x, x_s)$ y que $d_j(x, x_s) \geq d_M(x, x_s)$ por lo que la prueba anterior implica que también es denso en estos dos espacios. □

Usemos las técnicas que utilizamos en la prueba del teorema anterior, para obtener mas informacion de D.

Corolario 2.15. *Para todo* $\epsilon > 0$ *existe una sucesión* $0 = t_0 < t_1 < \ldots < t_n = T$ *tal que*

$$l(x[t_i, t_{i+1})) < \epsilon$$

para toda i*.*

Demostración. La sucesión de tiempos es $disc(x_s)$ tal como la construimos en el teorema. □

Corolario 2.16. *Toda función* $x \in D$ *es acotada.*

Demostración. Notemos que x_s toma una cantidad finita de valores, por lo tanto es acotada, además $||x_s - x|| < \epsilon$ por lo que

$$||x|| < |||x_s|| + \epsilon < \infty \, .$$

□

Corolario 2.17. *(D, M_1) y (D, J_1) son espacios separables..*

Demostración. Las funciones escalonadas con valores en los racionales y saltos en los racionales son numerables y densas en las funciones escalonadas, que son densas en D. □

2.4 Sobre la continuidad

En esta sección hablaremos de una peculiaridad del espacio D, metrizado con cualquiera de las métricas que hemos expuesto, la adición, vista como una función de $D \times D \to D$ no es continua.

Notemos que si la adición fuera continua, la suma de dos sucesiones convergentes convergería a la suma de los límites, para mostrar que la adición no es continua expongo el siguiente ejemplo:

Ejemplo 2.18. Tomemos $X_n = I_{[\frac{1}{2}-\frac{1}{n},1)}$ y $Y_n = -I_{[\frac{1}{2}+\frac{1}{n},1)}$, es claro que $X_n \to X$ y $Y_n \to Y$ donde $X = I_{[\frac{1}{2},1)}$ y $Y = -I_{[\frac{1}{2},1)}$.

Notemos que $X + Y = 0$, por lo tanto para que $X_n + Y_n$ convergiera a $X + Y$ tendria que tender a cero.

$$X_n + Y_n = I_{[\frac{1}{2}-\frac{1}{n},\frac{1}{2}+\frac{1}{n})} \, .$$

En cualquiera de las topologías esta sucesión claramente no converge a $X + Y$, por lo que se hace explícita la no continuidad.

2.5 Dominios infinitos

Ahora queremos ver qué pasa cuando el dominio ya no es finito, es decir, ahora estudiaremos la convergencia en $D(0, \infty)$.

La manera intuitiva de caracterizar la convergencia en $D(0, \infty)$ es pedir que para que converga en dicho espacio debe converger en $D[0, T]$ para todo $T \in \mathbb{R}$, sin embargo esto sería pedir demasiado como se muestra en el siguiente ejemplo:

Ejemplo 2.19. Sea $X_n = I_{[1+\frac{1}{n},\infty]}$ y sea $X = I_{[1,\infty]}$ desearíamos que X_n convergiera a X en $D(0,\infty)$, sin embargo X_n no converge en $D[0,T]$ para todo $T \in \Re$, en particular no converge para $D[0,1]$ pues en dicho espacio $X_n = 0$ y $X = I_{[1]}$ por lo tanto X_n no converge a X.

En general tenemos problemas en los puntos de discontinuidad de la función límite, por lo tanto diremos que una sucesion X_n converge a X en $D(0,\infty)$ si converge en $D[0,T]$ para todo $T \in \mathbb{R}$ tal que T es punto de continuidad de F.

Nos gustaría sintetizar la convergencia en todos los puntos de continuidad como la convergencia en alguna métrica, por esa razón definimos la métrica $D_{J_1\infty}$ de la siguiente manera:

Sean $x, y \in D(0,\infty)$ entonces:

$$D_{J_1\infty}(x,y) = \int_0^\infty e^{-t}[\min d_{J_1t}(x,y), 1]dt\,.$$

Podemos interpretar esa métrica como la esperanza de la función $d_{J_1t}(x,y)$ con t un tiempo de paro que se distribuye de manera exponencial y el uno es simplemente para que no se haga muy grande en algunos valores de t.

Veamos que en efecto es métrica:

Teorema 2.20. *$D_{J_1\infty}$ es métrica.*

Demostración. Sea x,y$\in D(0,\infty)$

• Como es la integral de cosas positivas es positiva.

• Si $x = y$ entonces d_{J_1} es cero y por lo tanto $D_{J_1\infty}(x,y) = 0$, si $D_{J_1\infty}(x,y) = 0$ entonces $d_{J_1T} = 0$ casi para todo t, de donde $x = y$ casi donde sea.

• Claramente $D_{J_1\infty}(x,y) = D_{J_1\infty}(y,x)$ pues esta métrica está en terminos de la métrica $J1$ que cumple el principio de simetría.

• Sólo queda probar la desigualdad del triángulo, ésta también se hereda de la métrica $J1$ en su versión finita como se muestra en las siguientes desigualdades:

$$\begin{aligned}
D_{J_1\infty}(x,y) &= \int_0^\infty e^{-t}[\min d_{J_1t}(x,y), 1]dt \\
&\leq \int_0^\infty e^{-t}[\min d_{J_1t}(x,z) + d_{J_1t}(z,y), 1]dt \\
&\leq \int_0^\infty e^{-t}[\min d_{J_1t}(x,z), 1]dt + \int_0^\infty e^{-t}[\min d_{J_1t}(z,y), 1]dt \\
&= D_{J_1\infty}(x,z) + D_{J_1\infty}(z,y)\,.
\end{aligned}$$

□

Ejemplo 2.21. Estudiaremos la convergencia de la siguiente sucesión en términos de $D_{J_1\infty}$. Sea $X_n = \sum_{i=1}^{\infty} iI_{[i-1+\frac{1}{n},i+\frac{1}{n})}$ y $X = \sum_{i=1}^{\infty} iI_{[i-1,i)}$, sea λ_n igual que en el ejemplo 3, para cada T fijo, entonces $D_{J_1\infty}(X, X_n) = \int_0^{\infty} e^{-t}[\text{mín}\, d_{J_1t}(X_n, X), 1]dt \leq \int_0^{\infty} \frac{e^{-t}}{n}dt = \frac{1}{n}$ que claramente tiende a cero.

Nota: la métrica $D_{J_1\infty}$ como la manejamos en este texto está sacada del libro de Ward Whitt, publicado en el 2002. Sin embargo no es la que siempre se ha usado, en el Ethier-Kurtz, de 1976 se usa:

$$D_{J_1\infty}(x,y) = \inf_{\lambda\in\Lambda}\left(\sup, \int_0^{\infty} e^{-t} \max_{u\in(0,t)} d(x(u)y(\lambda(u)))dt\right).$$

El supremo esencial actúa como el 1 en la versión de Whitt y la idea de la esperanza con un tiempo de paro que se distribuye exponencial es la misma. Sin embargo es mas difícil de utilizar, pues al no ser una constante el supremo esencial requiere más cuidado y tener el ínfimo afuera dificulta el cálculo directo.

Teorema 2.22. *Ahora veamos que las sucesiones convergen en $D(0,\infty)$ con la métrica $D_{J_1\infty}(x,y)$ si sólo si converge en la restricción a $D[0,T]$ con la métrica d_{J_1T} para todo $T \in \mathbb{R}$ tal que T es punto de continuidad de la función límite.*

Demostración. Sea X_n y X $\in D(0,\infty)$ tal que $D_{J_1\infty}(X, Xn) \to 0$ por definición de la métrica $J_1\infty$ sabemos que esto implica que $D_{J_1T}(x,y) \to 0$ para casi todo t, si para un t punto de continuidad no tendiera a cero la métrica D_{J_1T} entonces no lo haría para ninguno de los tiempos subsecuentes, por lo tanto $D_{J_1T}(x,y) \to 0$ para todo t punto de continuidad de X.

Supongamos que $D_{J_1T}(x,y) \to 0$ para todo t punto de continuidad de X, como X sólo puede tener una cantidad numerable de discontinuidades $d_{J_1t}(x,y) \to 0$ casi donde sea y por lo tanto $\int_0^{\infty} e^{-t}[\text{mín}\{d_{J_1t}(x,y), 1\}]dt \to 0$. □

2.6 Convergencia débil en el espacio de Skorohod

(Ver [8]) No es fácil probar directamente convergencia débil en el espacio D, expondremos brevemente la técnica que se usa para probar esto, primero definiremos distribucion finito dimensional:

Definición 2.23 (Convergencia en distribuciones finito dimensionales)**.** Decimos que una distribución finito dimensional es $(X_{n_{t_1}}, X_{n_{t_2}}, ..., X_{n_{t_n}})$ donde t_i son tiempos. Decimos que una sucesión de procesos estocastostico X_n converge en una distibución finito dimensionales a X un proceso estocastico si $(X_{n_{t_1}}, X_{n_{t_2}}, ..., X_{n_{t_n}}) \to (X_{t_1}, X_{t_2}, ..., X_{t_n})$ débilmente, para todo conjunto finito de tiempos $\{t_1, t_2...t_n\}$.

Notación: Si X_n converge débilmente a X lo denotaremos $X_n \Rightarrow X$.

El siguiente resultado se conoce como el teorema de Donsker y tiene muchas aplicaciones, por ejemplo la simulación computacional del movimiento Browniano.

Teorema 2.24. *Sea W la medida de Wiener, sean $\xi_1, \xi_2, \xi_3...$ variables aleatorias, independientes, idénticamente distribuidas, con esperanza cero y varianza $\sigma^2 < \infty$, sea $S_n = \xi_1 + \xi_2 + ... + \xi_n$ y sea*

$$X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]} \, ,$$

entonces la sucesión X_n converge débilmente a W sobre el espacio de Skorohod.

Sólo trabajamos la convergencia unidimensional. Para pasar a convergencia bidimensional tendriamos que utilizar técnicas mas complejas, demostrar, en general, convergencia finito dimensional es tedioso y complicado. Lo peor del caso es que incluso teniendo convergencia finito dimensional tendriamos que estudiar la tensión para asegurar la convergencia débil, como se verá en el siguiente teorema.

Sea t un tiempo fijo, entonces tenemos $X_n(t, \omega) = \frac{1}{\sigma\sqrt{n}} S_{[nt]}$. Calcularemos la función característica de X_n, denotaremos φ_Y la característica de W. Como $\xi_1, \xi_2, \xi_3, \ldots$ se distribuyen igual, denotaremos su característica φ_ξ.

Entonces

$$\begin{aligned}
\varphi_{X_n}(u) &= \varphi_{\frac{1}{\sigma\sqrt{n}} S_{[nt]}}(u) \\
&= \varphi_{S_{[nt]}}(\frac{u}{\sigma\sqrt{n}}) \\
&= \varphi_{\sum_{i=1}^{[nt]} \xi_i}(\frac{u}{\sigma\sqrt{n}}) \\
&= \prod_{1}^{[tn]} \varphi_\xi(\frac{u}{\sigma\sqrt{n}}) \\
&= (\varphi_\xi(\frac{u}{\sigma\sqrt{n}})^{[nt]} \, ..
\end{aligned}$$

Por otro lado recordemos que si tanto $E(\xi)$ como $E(\xi^2)$ son finitos, como es el caso por hipótesis, podemos asegurar que φ_ξ es clase C_2, entonces podemos aproximar la función característica φ_ξ por su serie de Taylor, alrededor del cero, calculemos la primera y segunda derivada,

$$\begin{aligned}\varphi'_\xi(u) &= E(i\xi e^{iu\xi})\\ \varphi''_\xi(u) &= E(-\xi^2 e^{iu\xi})\,,\end{aligned}$$

ahora valuemos dichas expresiones en $u = 0$,

$$\begin{aligned}\varphi'_\xi(u) &= E(i\xi e^{0}) = iE(\xi) = 0\\ \varphi''_\xi(u) &= E(-\xi^2 e^{0}) = E(\xi^2) = \sigma^2\,,\end{aligned}$$

con lo que se llega a lo siguiente:

$$\varphi_\xi(u) = 1 + 0 - \frac{\sigma^2 u^2}{2} + u^2 h(u)\,,$$

donde h es el error y cumple que $h(u)$ tiende a cero si u tiende a cero. Entonces:

$$\varphi_\xi(\frac{u}{\sigma\sqrt{n}}) = 1 - \frac{u^2}{2n} + \frac{u^2}{n\sigma^2}h(\frac{u}{\sigma\sqrt{n}})\,.$$

Ahora, regresando a la característica de x_n tenemos que

$$\begin{aligned}\varphi_{x_n} &= \varphi_\xi(\frac{u}{\sigma\sqrt{n}})^{[nt]}\\ &= \exp(\log(\varphi_\xi(\frac{u}{\sigma\sqrt{n}})^{[nt]}))\\ &= \exp([nt]\log\varphi_\xi(\frac{u}{\sigma\sqrt{n}}))\\ &= \exp([nt]\log(1 - \frac{u^2}{2n} + \frac{u^2}{n\sigma^2}h(\frac{u}{\sigma\sqrt{n}})))\,.\end{aligned}$$

Notemos entonces que $[nt]\log(1 - \frac{u^2}{2n} + \frac{u^2}{n\sigma^2}h(\frac{u}{\sigma\sqrt{n}})))$ cuando n tiende a infinito tiende a $-\frac{[nt]}{2n}u^2$ lo que claramente converge a $\frac{-tu^2}{2}$, de donde

$$\lim_{n\to\infty}\varphi_{x_n} = \lim_{n\to\infty}\exp([nt]\log(1 - \frac{u^2}{2n} + \frac{u^2}{n\sigma^2}h(\frac{u}{\sigma\sqrt{n}}))) = e^{\frac{-tu^2}{2}}\,.$$

Sin embargo nosotros sabemos que $e^{\frac{-u^2}{2}t}$ es la función característica de una normal $(0,t)$ y como tenemos un teorema que nos asegura que si las características convergen, entonces las variables aleatorias convergen en distribución, pues ya probamos lo que queríamos, $X_n(t)$ converge débilmente a una normal de esperanza cero y varianza t.

Se sugeriría consultar otras referencias, este es un resultado bien conocido que aparece en toda la bibliografía especializada.

Teorema 2.25. *Sean X y X_n procesos estocásticoscon trayectorias en D. Si para $X_n \to X$ en el sentido finito dimesional y además $\{P_n\}_{n\geq 1} = \{P \circ X_n^{-1}\}_{n\geq 1}$ es tenso, entonces $X_n \Rightarrow X$ débilmente en el espacio de Skorohod.*[1]

La idea es que si toda subsucesión de X_n tiene una subsucesión que converge débilmente a X, entonces X_n converge débilmente a X, se usa el hecho de que toda subsucesión tiene una subsucesión convergente por ser X_n tenso, luego se toma una arbitraria y como para todo t que cumple $P(\{x \in D : x(t) \neq x(t-)\}) = 0$ esa sucesión converge a X, se prueba que el conjunto de los tiempos que cumplen eso es densos, de donde la subsucesión a la que converge es igual a X en un denso, y de ahí se concluye que son iguales.

En esta parte cobra importancia el teorema de Prohorov que demostramos en el capitulo anterior, pues se convierte en algo de vital importancia conocer acerca de la tensión.
Es intuitivo pensar que si hay convergencia finito dimensional para todos los tiempos, deberia de haber convergencia débil sobre el espacio D, sin embargo no es así, el siguiente ejemplo nos mostrara como incluso en casos sencillos se puede dar convergencia finito dimensional, sin convergencia débil.

Ejemplo 2.26. Sea X, y X_n, tales que $P(X = 0) = 1$ mientras que $P(X_n = f_n) = 1$ con

$$f_n(t) = I_{[0,\frac{1}{n})}\,,$$

es decir, X_n es la medida de Dirac sobre f_n y X es la medida de Dirac sobre la función constante 0. Claramente f_n converge puntualmente a $I_{\{0\}}$, de donde las distribuciones finito dimensionales, para cualquier vector de tiempos, converge.

Sin embargo para toda n,

$$d_{j_n}(f, f_n) = 1\,,$$

Pero como nosotros estudiamos la distancia entre medidas de Dirac sabemos que eso implica que

$$P(||X - X_n|| = 1) = 1\,,$$

[1]Para una demostración refierase a *Convergence of Probability Measures* de Billingsley, pag. 124.

lo que muestra que no hay convergencia.

Esto nos está diciendo que las variables aleatorias no son tensas, (recuerde que una variable aleatoria sea tensa significa que las medidas imagen asociadas son tensas).

2.7 Oscilaciones

En esta última sección mencionaremos como se puede garantizar tensión en D, metrizado con J_1 y M_1. Con ese fin definimos las siguientes funciones que llamaremos oscilaciones. La razón del nombre es sugestiva por si misma, nos habla de como se aleja una funcion de un punto de su gráfica en una vecindad. Sea $x \in D$

Definición 2.27. $w_{ss}(x,t,h) = \text{ínf}_{\pi\in\Pi} \text{máx}_{v\in\pi} \sup_{r,s\in v} d(x(r),x(s))$.

Ahora definamos una función oscilación, la cual es parecida a la anterior, sólo que utiliza particiones. Sea Π_h el conjunto de todas las particiones de $[0,T)$ tales que, para todo intervalo $v=(a,b)\in\Pi_h$ se cumple que $b-a>h$, sea $x \in D$

Con esta función enunciamos el siguiente teorema que nos sera útil en el último capitulo, es una versión del teorema de Aldous que utiliza oscilaciones.

Teorema 2.28 (Importante)**.** *Sean X,X_1,$X_2,\ldots$ elementos aleatorios del espacio $D(R_+,R)$, entonces $X_n \Rightarrow X$ es equivalente a que se cumplan las siguientes 2 condiciones:*

- *convergen sus distribuciones finito dimensionales sobre algun denso T*
- $\text{lím}_{h\to 0} \text{lím}\sup_{n\to\infty} E[w_{ss}(X_n,t,h)] = 0$

En palabras, lo que estamos pidiendo en lugar de tensión, es que cuando la norma de las particiones se hace chica, el límite superior de la esperanza de la oscilacion de los procesos X_n tienda a cero. Es decir, las trayectorias que varian mucho en cachos chiquitos tendrán una probabilidad que tiende a cero. Pensar en el ejemplo donde no habia tensión nos ayudara a aceptar esta afirmación.[2].

[2]Para la demostración de esto remitase a *Probability and its applications* de Olav Kallenberg.

Capítulo 3

Procesos de Lévy

(Ver [1] y [7]) Un proceso de Lévy $\sigma(\xi_t : t \geq 0)$, es un proceso estocástico definido en un espacio de probabilidad $(\omega, \Im, P)$, con valores en $\mathbb{R}$, que cumple las siguientes características:

- Tiene trayectorias en D.
- Tiene incrementos independientes y estacionarios, es decir, para todo $s, t \geq 0$, $\xi_{t+s} - \xi_t$ es independiente de $\{\xi_u : u \leq t\}$ y tiene la misma distribucíon que ξ_s.
- $\xi_0 = 0$.

Los Procesos de Lévy son de gran utilidad pues tienen muchas propiedades que favorecen su estudio, sin que esto los haga una clase demasiado pequeña, por eso los Procesos de Lévy tienen aplicaciones muy diversas y son de gran interés.

Sin lugar a dudas el proceso de Lévy más conocido es el movimiento Browniano, nosotros tenemos la ventaja de haber expuesto el teorema de Donsker, pues pensar al movimiento Browniano como límite de caminatas aleatorias nos ayudará a entender las propiedades de los Procesos de Lévy de forma más intuitiva.

Una propiedad muy importante de los Procesos de Lévy, es que tienen la propiedad de Markov. Nosotros demostraremos que los Procesos de Lévy cumplen la propiedad de Markov débil en el siguiente teorema.

Teorema 3.1. *Si $(X_t)_{t\geq 0}$ es un proceso de Lévy y $h \geq 0$, el proceso $(X_{t+h} - X_h)_{t\geq 0}$ es Proceso de Lévy y*

$$Y_t = X_{t+h} - X_h \doteq X_t \,.$$

La prueba es muy simple, es claro que

$$Y_0 = X_h - X_h = 0 \,.$$

Además como la suma de trayectorias Cadlag es Cadlag, tenemos la primera propiedad de Procesos de Lévy. Por otro lado tiene incrementos independientes porque X es Proceso de Lévy, así que sólo falta ver que sea estacionario, es decir que para

$$h, t \geq 0,\, Y_{t+s} - Y_h = Y_t \text{ en distribución}\,.$$

Lo que es cierto puesto que

$$Y_{t+s} - Y_t = (X_{t+s+h} - X_h) - (X_{t+h} - X_h)\,,$$

pero como $(X_t)_{t\geq 0}$ es proceso de Lévy tenemos que los incrementos son estacionarios, lo que significa que

$$X_{t+s+h} - X_h \doteq X_{t+s}\,, X_{t+h} - X_h \doteq X_t\,;$$

pero por definición

$$X_{t+s} - X_t \doteq Y_s\,,$$

lo que termina la prueba.

Definición 3.2. el tiempo de paro es una variable aleatoria no negativa T, tal que

$$\{T \leq t\} \in \Im_t\,.$$

Lo que nosotos vamos a utilizar en la última sección es un resultado mas fuerte, conocido como propiedad de Markov fuerte y que enunciamos en el siguiente teorema.

Teorema 3.3. *Si $(X_t)_{t\geq 0}$ es un Proceso de Lévy y τ es un tiempo de paro finito con respecto a la filtración de $(X_t)_{t\geq 0}$, entonces el proceso $(X_{t+\tau} - X_\tau)_{t\geq 0}$ es proceso de Lévy y*

$$Y_t = X_{t+\tau} - X_\tau \doteq X_t\,.$$

Esto se prueba primero para tiempos de paro discretos y luego aproximando cualquier tiempo de paro por una sucesión de tiempos de paro discretos. Los detalles se pueden consultar en cualquier libro de la bibliografía.

3.1 Medidas infinitamente divisibles

Decimos que una variable aleatoria real es infinitamente divisible si para toda n, la podemos expresar como suma de n variables aleatorias independientes e idénticamente distribuidas.

Es decir, si ξ es infinitamente divisible, para toda n, existe $(z_i)_{i=1}^n$ tal que

$$\xi = \sum_1^n z_i, \text{ con } (z_i)_{i=1}^n \text{ independientes e idénticamente distribuidas}.$$

Las variables aleatorias infinitamente divisibles tienen importancia y aparecen en muchas aplicaciones.Para nosotros lo importante es el siguiente teorema.

Teorema 3.4. *Si $(\xi_t)_{t\geq 0}$ es un proceso de Lévy ξ_t, para toda $t > 0$ la ley de ξ_t es infinitamente divisible. Además existe $\psi\colon \mathbb{R} \to C$ tal que*

$$\varphi_{\xi_t}(\lambda) = E(\exp(i\lambda\xi_t)) = \exp(-t\psi(\lambda))\,,$$

donde $\psi\colon \mathbb{R} \to C$ es conocida como el exponente característico del proceso ξ.

NOTA: el exponente caracteristico es un objeto muy importante, pués caracteriza al proceso de Lévy.

Sea $(\xi_t)_{t\geq 0}$ un proceso de Lévy, la primera parte es muy simple, basta utilizar la propiedad de incrementos independientes y estacionarios para escribir

$$\xi_t = \xi_{\frac{t}{n}} - \xi_{\frac{t}{n}} + \xi_{\frac{2t}{n}} - \xi_{\frac{2t}{n}} + \cdots + \xi_{\frac{nt}{n}}\,,$$

la igualdad es muy clara, pues es una suma telescópica. Luego podemos agrupar,

$$\xi_t = \xi_{\frac{t}{n}} + (\xi_{\frac{2t}{n}} - \xi_{\frac{t}{n}}) + \cdots + (\xi_{\frac{nt}{n}} - \xi_{\frac{(n-1)t}{n}})\,,$$

por la propiedad de incrementos independientes y estacionarios, las variables aleatorias dentro de cada paréntesis son independientes e idénticamente distribuidas, así que en distribución

$$\xi_t = \sum_1^n z_j\,,$$

donde $z_j \doteq \xi_{\frac{jt}{n}} - \xi_{\frac{(j-1)t}{n}}$.

Antes de continuar con el teorema, recordaremos a través de un lema algunas propiedades de la función característica.

Lema 3.5. *Si Y es una variable aleatoria infinitamente divisible su función característica $\varphi_Y : \mathbb{R} \to C$ cumple que:*

- *para toda $n \in N$ existen $(z_i)_{i=1}^n$ variables aleatorias independientes e identicamente distribuidas tales que $\varphi_Y(\lambda) = (\varphi_{z_i}(\lambda)^n$;*
- *φ_Y nunca se anula;*
- *existe $\psi \colon \mathbb{R} \to C$ tal que $\log \varphi_Y(\lambda) = -\psi(\lambda)$.*

Demostración. La prueba es la siguiente, sea $n \in N$ y $(z_j)_{j=1}^n$ las variables aleatorias independientes idénticamente distribuidas tales que

$$\varphi_Y(\lambda) = E(\exp(i\lambda Y)) = E(\exp(i\lambda \sum_{j=1}^n z_j))$$
$$E(\Pi_{j=1}^n exp(i\lambda z_j))\,;$$

Como son identicamente distribuidas, basta tomar $z = z_1$ para tener

$$\Pi_{j=1}^n E(\exp(i\lambda z_j)) = E(\exp(i\lambda z_1))^n = (\varphi_z(\lambda))^n\,.$$

Por otro lado notemos que el lema anterior nos asegura que la función característica de Y nunca es cero, esto no lo probaremos rigurosamente, pero la idea heurística es la siguiente: Pensemos en procesos de Lévy, sea $(\xi_t)_{t\geq 0}$ un proceso de Lévy, entonces

$$\varphi_{\xi_t}(\lambda) = \varphi_{\xi_{\frac{t}{n}}}(\lambda)^n\,,$$

para cualquier n. Supongamos que se anula, tendríamos

$$0 = \varphi_{\xi_t}(\lambda) = \varphi(\xi_{\frac{t}{n}})^n\,.$$

Lo que implicaría que la función característica es cero para todo $\xi_{\frac{t}{n}}$, para todo proceso de Lévy, pero esto no es verdad, la contradicción viene de suponer que $0 = \varphi_{\xi_t}(\lambda)$.

Podemos asegurar que el logaritmo de la función característica, $\log(\varphi(Y))$, está bien definido y tiene una rama continua, pues no se anula. Esa rama será el exponente característico[1],

$$\psi = -\log(\varphi_Y(\lambda))\,,$$

luego es claro que

$$\varphi_Y(\lambda) = \exp(-\psi(\lambda)) = \exp(\log(\varphi_Y(\lambda)))\,,$$

lo que termina la prueba del lema. □

[1] Una prueba completa de estos dos últimos hechos puede revisarse en [1]

El lema anterior implica que el teorema se cumple para $t = 1$, es decir que

$$\varphi_{\xi_1}(\lambda) = E(\exp(i\lambda\xi_1)) = \exp(-\psi(\lambda)),$$

sólo nos falta ver que se cumple para la infinidad no numerable de reales positivos distintos de 1, lo hacemos de la siguiente forma:

Para cualquier natural n tenemos que

$$\xi_n = \xi_1 + (\xi_2 - \xi_1) + ...(\xi_n - \xi_{n-1}),$$

y así

$$\varphi_{\xi_n}(\lambda) = (\varphi_{\xi_1}(\lambda))^n = \exp(-n\psi(\lambda)),$$

por lo tanto en los naturales también se cumple. También podemos ver que para todo natural k

$$E(\exp \lambda i\xi_1) = (E(\exp \lambda i\xi_{\frac{1}{k}}))^k = \exp(-\psi(\lambda)),$$

de donde concluímos que

$$E(\exp \lambda i\xi_{\frac{1}{k}}) = \exp(-\frac{1}{k}\psi(\lambda)).$$

De estos dos casos fácilmente concluimos, combinando los métodos, que se cumple para todo racional, es decir

$$E(\exp i\lambda\xi_{\frac{m}{k}}) = \exp(-\frac{m}{k}\psi(\lambda)).$$

Tenemos entonces el resultado para todos los racionales, para llegar a los irracionales simplemente tomamos una sucesión decreciente de racionales que converga a un irracional; por continuidad de las trayectorias por la derecha, concluimos que tambien para los irracionales es válido este teorema, lo que termina la prueba.

Más adelante calcularemos los exponentes característicos de los más relevantes procesos de Lévy.

3.2 Martingalas

En un espacio de probabilidad $(\Omega, \Im, P)$, en el que se tiene una filtación $(\Im_t)_{t\geq 0}$, $\Im_t \subset \Im,\ t \geq 0$, se dice que el proceso estocastico $(M_t)_{t\geq 0}$ definido en $(\Omega, \Im, P)$ es una martingala si cumple

1. $M_t \in L_1(\Omega, \Im, P),\ t \geq 0$,

2. $E(M_t|\Im_s) = M_s$ siempre que $s \leq t$.

Sobre las martingalas tenemos los siguientes 3 importantes teoremas

Teorema 3.6 (Desigualdad maximal de Doob). *Para todo $t > 0$ se cumple que*

$$E(\sup\{|M_s|^2 : 0 \geq s \geq t\}) \leq 4E(|M_t|^2)\,.$$

Teorema 3.7 (Teorema del paro opcional). *Sea T un tiempo de paro finito casi seguramente, entonces:*

1. *el proceso $M_{\text{mín}(t,T)}$ es una martingala,*

2. *además si M es uniformemente integrable*

$$E(M_T) = E(M_0)\,.$$

Teorema 3.8 (Convergencia de martingalas). *Si M es uniformemente integrable, $\text{lím}_{t\to\infty} M_t = M_\infty$ existe casi seguramente y*

$$M_t = E(M_\infty|\Im_t)\,.$$

3.3 Ejemplos relevantes de procesos de Lévy y martingalas asociadas

En esta seccion se daran algunos ejemplos importantes de procesos de Lévy.

3.3.1 Proceso Poisson

El proceso Poisson $\{N_t\}_{t\geq 0}$ de intensidad $c > 0$, es un proceso de Lévy creciente, tal que N_t se distribuye Poisson de parámetro ct.

Construiremos dicho proceso de la siguiente manera: tomamos una sucesion de variables aleatorias $\tau_1, \tau_2, \ldots$ tales que

$$P(\tau_i > s) = e^{-cs}\,.$$

Es decir, τ_i se distribuye exponencial de parametro c, definimos $S_n = \sum_1^n \tau_i$, sabemos que la suma de variables aleatorias exponenciales es una variable aleatoria con distribución Gamma, luego entonces, para $s > 0$

$$P(S_n \in ds) = s^{n-1}e^{-cs}ds\,.$$

Proponemos

$$N_t = \sup\{n \in N : S_n \leq t\}$$

veamos que en efecto es el proceso Poisson de intensidad c,

$$\begin{aligned} P(N_t = k) &= P(\sup\{n \in N : S_n \leq t\} = k) \\ &= P(S_k \leq t, S_{k+1} > t) \\ &= \int_0^t \frac{c^k}{(k-1)!} s^{k-1} e^{-cs} e^{-c(t-s)} ds \\ &= \frac{c^k}{(k-1)!} e^{-ct} \int_0^\infty s^{k-1} \frac{t^k}{t^k} = e^{-ct}(ct)^k/(k!)\,. \end{aligned}$$

Esto prueba que $P(N_t = k) = e^{-ct}(ct)^k/(k!)$, es decir N_t se distribuye Poisson con parametro ct.

Se sabe que el proceso Poisson tiene incrementos independientes e idénticamente distribuidos.

Calculemos el exponente característico, tenemos que

$$\varphi_{N_1}(\lambda) = E(\exp \lambda i N_1)\,,$$

por definición, luego

$$E(\exp \lambda i N_t) = \sum_{k=0}^{\infty} e^{i\lambda k} e^{-c} c^k / k!\,,$$

también por definición, ahora

$$\sum_{k=0}^{\infty} e^{i\lambda k} e^{-c} c^k / k! = e^{-c} \sum_{k=0}^{\infty} (ce^{i\lambda})^k / k!\,,$$

simplemente agrupando, pero

$$e^{-c} \sum_{k=0}^{\infty} (ce^{i\lambda})^k / k! = e^{-c} e^{c(e^i\lambda)}\,,$$

pues conocemos la expansión en serie de Taylor de la exponencial. Finalmente tenemos que

$$e^{-c} e^{c(e^i\lambda)} = \exp(c(e^{i\lambda} - 1))\,,$$

de donde en conclusión

$$\psi(\lambda) = c(1 - e^{i\lambda})\,.$$

Tenemos 3 martingalas de interés asociadas al proceso Poisson

Teorema 3.9. *Si N_t es proceso de Poisson, los siguientes 3 procesos son martingalas:*

1. $M_t = N_t - ct$,
2. $M_t^2 - ct$,
3. $\xi_t^q = \exp(-qN_t + ct(1 - e^{-q}))$ *donde* $q > 0$.

Demostración. 1. Usemos la propiedad de incrementos independientes de la siguiente manera: Sea $s < t$, entonces

$$E[M_t - M_s|\Im_s] = E[N_t - ct - N_s + cs|\Im_s] = E[N_t - N_s|\Im_s] - c(t-s)\,,$$

pero por incrementos independientes tenemos que

$$E[N_t - N_s|\Im_s] - c(t-s) = E[N_t - N_s] - c(t-s)\,,$$

pero sabemos que N_t tiene distribución Poisson de parametro ck, entonces

$$E[N_t - N_s] - c(t-s) = c(t-s) - c(t-s) = 0\,,$$

por lo tanto M_t es martingala.

2. Primero desarrollemos la martingala, en términos de M_t y usemos independencia de incrementos. Tenemos que $(M_t - M_s)^2 + 2M_tM_s = M_t^2 + M_s^2$, por lo que

$$\begin{aligned}
E[M_t^2 - ct - M_s^2 - cs|\Im_s] &= E[M_t^2 - M_s^2|\Im_s] - c(t-s) \\
&= E[(M_t - M_s)^2 + 2M_tM_s + 2M_s^2|\Im_s] - c(t-s) \\
&= E[(M_t - M_s)^2|] + 2E[M_s(M_t - M_s)|\Im_s] - c(t-s)\,.
\end{aligned}$$

Analizemos por separado el segundo sumando:

$$2E[M_s(M_t - M_s)|\Im_s] = 2M_sE[M_t - M_s|\Im_s]\,,$$

por incrementos independientes, y porque M_s es $\Im_s$ medible. Pero en el inciso anterior probamos que

$$E[M_t - M_s|\Im_s] = 0\,,$$

por lo tanto el segundo sumando es cero. Para mostrar que es martingala, basta ver que el primer sumando es $c(t-s)$, eso se muestra de la siguiente manera: tenemos que

$$M_t^2 = (N_t - ct)^2 = N_t^2 - 2ctN_t + c^2t^2\,,$$

entonces

$$\begin{aligned} E[(M_t - M_s)^2] &= E[(N_t - N_s)^2 - 2c(t-s)(N_t - N_s) + c^2(t-s)^2] \\ &= E[(N_t - N_s)^2] - 2c(t-s)E[N_t - N_s] + c^2(t-s)^2 \, , \end{aligned}$$

pero

$$E[N_t^2 - N_s^2] = \mathrm{var}(N_{t-s}) + E(N_{t-s})^2 = c(t-s) + c^2(t-s)^2 \, ,$$

y claramente

$$2c(t-s)E[N_t - N_s] = 2c^2(t-s)^2 \, .$$

Sustituyendo estas dos ecuaciones tenemos que en efecto

$$E[(M_t - M_s)^2] = c(t-s) \, ,$$

lo que nos asegura que

$$E[M_t^2 - ct - M_s^2 - cs|\Im_s] = 0 \, .$$

Lo que prueba que $M_t^2 - ct$ es martingala.

3. Sea $s < t$, tenemos las siguientes igualdades

$$E[\xi_t^q|\Im_s] = E[\exp(-qN_t + ct(1-e^{-q}))|\Im_s],$$

por definicion, luego

$$E[\exp(-qN_t + ct(1-e^{-q}))|\Im_s] = E[\exp(-q(N_t - N_s + N_s) + c(t-s+s)(1-e^{-q}))|\Im_s] \, ,$$

simplemente sumando ceros, pero esto lo podemos agrupar para aprovechar los incrementos independientes

$$E[\exp(-q(N_t - N_s + N_s) + c(t-s+s)(1-e^{-q}))|\Im_s]$$

$$= E[\exp(-q(N_t - N_s) + c(t-s)(1-e^{-q}))\exp(-q(N_s) + c(s)(1-e^{-q}))|\Im_s] \, ;$$

como $\exp(-q(N_s) + c(s)(1-e^{-q}))$ es $\Im_s-$medible tenemos la siguiente igualdad

$$E[\exp(-q(N_t - N_s) + c(t-s)(1-e^{-q}))\exp(-q(N_s) + c(s)(1-e^{-q}))|\Im_s]$$

$$= \exp(-q(N_s) + c(s)(1-e^{-q}))E[\exp(-q(N_t - N_s) + c(t-s)(1-e^{-q}))|\Im_s] \, .$$

Fijémonos en lo que quedó dentro de la esperanza. Por independencia de incrementos podemos quitar la sigma algebra con la que estamos condicionando, mientras que por estacionalidad de incrementos podemos calcular explícitamente esta esperanza

$$
\begin{aligned}
& E[\exp(-q(N_t - N_s) + c(t-s)(1-e^{-q}))|\Im_s] \\
&= E[\exp(-q(N_t - N_s) + c(t-s)(1-e^{-q}))] = \\
&= \sum_{k=0}^{\infty} \exp(-qk + c(t-s)(1-e^{-q})\frac{(c(t-s))^k e^{-c(t-s)}}{k!} \\
&= \exp(c(t-s)(1-e^{-q}-1)\sum_{k=0}^{\infty} \frac{e^{-q}(c(t-s))^k}{k!} \\
&= e^{-c(t-s)e^{-q}} e^{c(t-s)e^{-q}} = e^0 = 1\,,
\end{aligned}
$$

de donde si substituimos obtenemos que

$$
\begin{aligned}
& E[\xi_t^q|\Im_s] = \\
&= \exp(-q(N_s) + c(s)(1-e^{-q}))E[\exp(-q(N_t - N_s) + c(t-s)(1-e^{-q}))|\Im_s] \\
&= \exp(-q(N_s) + c(s)(1-e^{-q})) = \xi_s\,.
\end{aligned}
$$

□

3.3.2 Movimiento Browniano

Un movimiento Browniano $(B_t)_{t\geq 0}$ es un proceso de Lévy continuo casi seguramente, tal que

$$B_t \sim N(0,t)\,.$$

Es decir, para todo tiempo t, la variable B_t tiene una distribución normal de esperanza 0 y varianza t, tiene incrementos independientes y estacionarios y es cero en cero.

El movimiento Browniano es uno de los objetos matemáticos más útiles en la práctica, el teorema de Donsker nos asegura que e muchos fenómenos se pueden simular con un movimiento Browniano. En finanzas el movimiento Browniano y sus procesos asociados modelan el comportamiento de tantas cosas que no se puede concebir la matemática financiera sin movimiento Browniano. Además este proceso tiene muchas interesantes propiedades que enunciaremos en esta subsección, las cuales lo hacen, en muchos casos, una rica fuente de respuestas.

Calculemos su exponente característico, nos fijamos en B_1 y tenemos que:

$$\varphi_{B_1}(\lambda) = E(\exp \lambda i B_1),$$

como B_1 se distribuye normal $(0,1)$

$$E(\exp \lambda i B_1) = \int_{-\infty}^{\infty} e^{i\lambda x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx \, .$$

Como conocemos la función compleja $e^i x = \cos x + i \operatorname{sen} x$ tenemos que

$$\int_{-\infty}^{\infty} e^{i\lambda x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx = \int_{-\infty}^{\infty} \cos \lambda x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx + i \int_{-\infty}^{\infty} \operatorname{sen} \lambda x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx \, .$$

Recordemos que $\operatorname{sen}(\lambda x e^{\frac{x^2}{2}})$ es una función impar, lo que implica que

$$i \int_{-\infty}^{\infty} \operatorname{sen} \lambda x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx = 0 \, ,$$

por lo que sólo debemos fijarnos en el primer sumando, es decir

$$\varphi_{B_1}(\lambda) = \int_{-\infty}^{\infty} \cos \lambda x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx \, ,$$

como $E(|B_1|) < \infty$ sabemos que la función característica tiene primera derivada y es continua, entonces podemos derivar de ambos lados la ecuación anterior, además por convergencia dominada podemos derivar abajo del signo de integral, entonces

$$\varphi\text{'}_{B_1}(\lambda) = \int_{-\infty}^{\infty} -x \operatorname{sen} \lambda x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx \, .$$

Ahora tomamos $v = \operatorname{sen}(\lambda x)$, $u = x e^{\frac{-x^2}{2}}$ y usamos la fórmula de integración por partes

$$udv = uv - \int vdu \, ,$$

además de argumentos de funciones impares para obtener

$$\varphi\text{'}_{B_1}(\lambda) = \int_{-\infty}^{\infty} -\lambda \cos \lambda x \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx \, ,$$

o lo que es equivalente

$$\varphi\text{'}_{B_1}(\lambda) = -\lambda \varphi_{B_1}(\lambda) \, ,$$

de donde obtenemos la bien conocida ecuación diferencial

$$\frac{\varphi'_{B_1}(\lambda)}{\varphi_{B_1}(\lambda)} = -\lambda\,,$$

que implica

$$\ln|\varphi_{B_1}(\lambda) = 1| = -\frac{\lambda^2}{2} + c\,,$$

lo que nos lleva, aplicando exponencial de ambos lados a

$$\varphi_{B_1}(\lambda) = e^{-c}e^{-\frac{\lambda^2}{2}}\,,$$

tomamos c para que $\varphi_{B_1}(0) = 1$ y concluímos

$$\varphi_{B_1}(\lambda) = e^{-\frac{\lambda^2}{2}}\,.$$

En consecuencia el exponente caracteristico es $\psi(\lambda) = \frac{\lambda^2}{2}$.

Enunciaremos algunos procesos de interés asociados al movimiento Browniano.

1. Para cualquier $x \in R$ el proceso $X_t^x = x + B_t$ es conocido como el movimiento Browniano empezado en x, su distribucion es claramente normal (x, t), es decir

$$P[X_t^x \in A] = \frac{1}{\sqrt{2\pi t}} \int_A e^{-(y-x)^2/2t}\,.$$

2. Proceso

$$X(t) = B(t) + \mu t\,.$$

   Es conocido como movimiento Browniano con deriva μ. Una manera alterna de definirlo es decir que un movimiento Browniano con deriva es un proceso estocástico que cumple las siguientes propiedades:

   - $X(0) = 0$,
   - $\{X(t) : t \geq 0\}$ tiene incrementos independientes y estacionarios,
   - $X(t)$ se distribuye Normal $(\mu t, t)$;

   y tambien se puede ver como limite débil de caminatas aleatorias simples.

3. Si $B_t^1, B_t^2, \ldots, B_t^d$, son d copias independientes del movimiento Browniano, definimos
$$X_t^d = (B_t^1, B_t^2 ... B_t^d)\,,$$
un movimiento Browniano d dimensional. Este proceso resulta interesante entre otras cosas por sus propiedades de recurrencia o transitividad para algunos valores de d, también es interesante preguntarse por la probabilidad de que alcance alguna barrera en vez de otra.

Ahora veremos algunos procesos asociados al movimiento Browniano.

1. El proceso (t, X_t^d) es conocido como el proceso de la ecuación de calor en $R^+ \times R^d$.

2. Algunas veces nos interesa conocer los valores máximos del movimiento Browniano en $\Re$, entonces definimos
$$S_t = sup\{B_s : 0 \leq s \leq t\}\,.$$
Cabe destacar que el supremo sobre todos los posibles valores de s, es el mismo que el supremo sobre todos los racionales, por continuidad del movimiento Browniano.

3. El inverso del movimiento Browniano consiste en aplicar la función inversa generalizada
$$B_t^{-1} = inf\{s : B_s \geq t\}\,,$$
este proceso es creciente y Cadlag.

4. Definimos el proceso de tiempo de ocupacion un Boreliano A, como
$$O_t = \int_0^t I_A(B_s) ds\,.$$
Este proceso tiene sentido pues, por continuidad del movimiento Browniano, la función
$$(s, \omega) \to I_A(B_s(\omega))\,,$$
es medible en $(\Omega \times R^+)$ con sigma álgebra producto$B(\mathbb{R}) \otimes \Im$.

5. Definimos el Browniano Geométrico como
$$Y(t) = e^{B_t}\,.$$

Conocer la esperanza y la varianza del Browniano Geoḿetrico es cosa fácil dado que conocemos la función generadora de momentos del movimiento Browniano, es decir como

$$E[e^{\lambda B_t}] = e^{\frac{t\lambda^2}{2}} \,,$$

claramente

$$E[Y(t)] = E[e^{B_t}] = e^{\frac{t}{2}} \,,$$

mientras que

$$\mathrm{Var}(Y(t)) = E(Y(t)^2) - E(Y(t))^2 = E[e^{2B_t}] - E[e^{B_t}]^2 = e^{2t} - e^t \,.$$

Ahora expondremos algunas martingalas de interés asociadas al movimiento Browniano.

1. Sin duda la martingala mas importante asociada al movimiento Browniano, es el mismo movimiento Browniano. Probemos que en efecto es martingala. Sea $t > s$

$$E[B_t|\Im_s] = E[B_t - B_s + B_s|\Im_s] = E[B_t - B_s|\Im_s] + [B_s|\Im_s]$$

por linealidad, analizemos el primer sumando por separado

$$E[B_t - B_s|\Im_s] = E[B_t - B_s] = E[B_t] - E[B_s] = 0 - 0 = 0$$

donde la primera igualdad es por incrementos independientes, la segunda por linealidad y la tercera por construcción del Browniano. Entonces tenemos que

$$E[B_t|\Im_s] = E[B_s|\Im_s] \,,$$

sin embargo como B_s es $\Im_s-$medible, tenemos que

$$E[B_s|\Im_s] = B_s \,,$$

de donde

$$E[B_t|\Im_s] = B_s \,,$$

y por lo tanto B_t es una martingala.

2. La siguiente martingala es caso particular de una familia de martingalas de gran importancia en el cálculo estocástico, que son las martingalas al cuadrado menos su variación cuadrática, la variación cuadrática de B_t es t, por lo que

$$X_t = B_t^2 - t$$

es martingala.

Probemos este hecho,

$$E[X_t|\Im_s] = E[B_t^2|\Im_s] - t\,,$$

por linealidad de la esperanza, entonces sólo nos fijaremos en el primer sumando provisionalmente

$$E[B_t^2|\Im_s] = E[((B_t - B_s) + B_t)^2|\Im_s]\,,$$

desarrollando el cuadrado tenemos

$$\begin{aligned} E[((B_t - B_s) + B_s)^2|\Im_s] &= E[(B_t - B_s)^2 + 2(B_t - B_s)(B_s) + B_s^2|\Im_s] \\ &= E[(B_t - B_s)^2|\Im_s] + E[2(B_t - B_s)(B_s)|\Im_s] \\ &\quad + E[B_s^2|\Im_s]\,, \end{aligned}$$

donde la última igualdad es por linealidad. Analizemos cada sumando por separado,

$$E[(B_t - B_s)^2|\Im_s] = E[(B_t - B_s)^2]\,,$$

por incrementos independientes, pero por incrementos idénticamente distribuidos sabemos que $B_t - B_s$ se distribuye Normal $(0, t - s)$, por tener esperanza cero, sabemos que

$$\mathrm{Var}[B_t - B_s] = E[(B_t - B_s)^2] = t - s\,.$$

El segundo sumando

$$E[2(B_t - B_s)(B_s)|\Im_s] = 2E[(B_t - B_s)|\Im_s]E[(B_s)|\Im_s]\,,$$

por incrementos independientes, además

$$E[(B_s)|\Im_s] = B_s < \infty \qquad \text{c.s.}\,,$$

donde la igualdad es por ser B_s medible y la desigualdad por que B_s se distribuye normal. Mientras que

$$E[(B_t - B_s)|\Im_s] = E[(B_t - B_s)] = E[B_t] + E[B_s] = 0\,,$$

de donde

$$E[2(B_t - B_s)(B_s)|\Im_s] = 0\,.$$

Por último el tercer sumando se analiza en un renglón, por ser B_s $\Im_s$-medible

$$E[B_s^2|\Im_s] = B_s^2\,.$$

Uniendo todo esto tenemos que:

$$E[B_t^2|\Im_s] = B_s^2 + t - s\,,$$

por lo tanto

$$E[X_t|\Im_s] = B_s^2 + t - s - t = X_s\,.$$

Lo que nos permite concluir que X_t es martingala.

3. La martingala exponencial asociada al Browniano, se define como

$$Y(t) = \exp\{cB(t) - c^2\frac{t}{2}\}\,.$$

Enunciaremos algunas propiedades del movimiento Browniano.

- Homogenidad.

  Para todo $s > 0$ el proceso estocástico $Y_t = B_{t+s} - B_s$ es movimiento Browniano.

  Demostración. Claramente $Y_t = 0$ pues $Y_0 = B_t - B_t = 0$, tiene incrementos independientes pues $Y_t - Y_q = B_{t+s} - B_s - (B_{q+s} - B_s) = B_{t+s} - (B_{q+s}$ que es un incremento de B_s y como B_s es movimiento Browniano es independiente de cualquier otro incremento ajeno. La continuidad casi segura se hereda trivialmente. Se distribuye Normal por que el movimiento Browniano tiene incrementos estacionarios y $Y_t = B_{t+s} - B_s$ es un incremento de longitud t. Por último tiene incrementos estacionarios, pues $Y_t - Y_q = B_{t+s} - B_s - (B_{q+s} - B_s) = B_{t+s} - B_{q+s}$, $q < t$ es un incremento de longitud $t - q$, que por estacionariedad de B_t se distribuye Normal $(0, t - q)$. Lo que termina la prueba. $\square$

- Simetría.

  El proceso $-B_t$ es movimiento Browniano.

Demostración. La prueba se deriva trivialmente de la simetria de la distribucion Normal. □

- Escalamiento.

  Para todo $c > 0$ el proceso $Y_t = cB_{\frac{t}{c^2}}$ es movimiento Browniano.

  Demostración. La unica propiedad que requiere prueba es que Y_t se distribuye Normal$(0, t)$, es claro que Y_t se distribuye Normal, pero veamos que en efecto los parametros son los que queremos;

  $$E(Y_t) = E(cB_{\frac{t}{c^2}}) = cE(B_{\frac{t}{c^2}}) = 0\,,$$

  donde la última igualdad se da porque $B_{\frac{t}{c^2}} \sim N(0, \frac{t}{c^2})$.

  $$\mathrm{Var}(Y_t) = E(cB^2_{\frac{t}{c^2}}) = c^2E(B^2_{\frac{t}{c^2}}) = t\,,$$

  donde de nuevo la última igualdad es porque $B_{\frac{t}{c^2}} \sim N(0, \frac{t}{c^2})$. □

- Tiempo invertido.

  El proceso X_t definido como $X_0 = 0$, $X_s = sB_{\frac{1}{s}}$ es un movimiento Browniano.

  Demostración. Para ver que $X_s \to 0$ cuando $s \to 0$ se requieren herramientas mas avanzadas, por lo que no lo probaremos, sin embargo con técnicas similares a las que usamos para el escalonamiento mostraremos que tiene distribucion Normal$(0, t)$. Es claro que $E(X_s) = 0$, además

  $$\mathrm{Var}(X_s) = E(X_s^2) = s^2E(B_{\frac{1}{s}}) = s\,.$$

  Esto prueba que en efecto de distribuye Normal$(0, s)$. □

- Recurrencia y transitoriedad del movimiento Browniano.

  El movimiento Browniano 1−dimensional cumple que todos sus puntos son recurrentes.

  En dos dimensiones el cero es recurente por vecindades, es decir, para cualquier vecindad del cero el movimiento Browniano visitará una infinidad de veces esa vecindad con probabilidad 1.

Para $n \geq 3$, el movimiento Browniano $n-$dimensonal tiene todos sus puntos transitorios.

La prueba de esto requiere conceptos mas profundos de Martingalas, por lo que no se desarrollará en el presente trabajo.

3.3.3 Proceso Poisson compuesto

Sea $\xi_1, \xi_2, \ldots$ una sucesión de variables aleatorias independientes e idénticamente distribuidas, con distribución v, en $\mathbb{R}^d - \{0\}$. Sea N_t un proceso Poisson de parametro c, independiente de ξ_i, para todo i, el proceso

$$\sum_{1}^{N_t} \xi_i$$

es el proceso Poisson compuesto.

Su exponente característico se deduce a continuación,

$$\varphi_X(\lambda) = E\left(\exp i\lambda \sum_{i=1}^{N_t} \xi_i\right)$$

por definicion, pero

$$E\left(\exp i\lambda \sum_{i=1}^{N_t} \xi_i\right) = E\left(\sum_{n=0}^{\infty} I_{N_t=n} \exp i\lambda \sum_{i=1}^{n} \xi_i\right),$$

pues $\sum_{n=0}^{\infty} I_{N_t=n} = \Omega$, además es una partición disjunta, entonces

$$E\left(\sum_{k=0}^{\infty} I_{N_t=n} \exp i\lambda \sum_{i=1}^{n} \xi_i\right) = \sum_{n=0}^{\infty} E\left(\exp i\lambda \sum_{i=1}^{n} \xi_i\right) P(N_t = n)$$

esto se debe a que N_t es independiente de ξ_i, para todo i. Pero por ser N_t proceso Poisson y la suma en el coeficiente de una exponencial, es igual a la multiplicación de exponenciales con los sumandos como exponentes por lo que esto último es igual a

$$\sum_{n=0}^{\infty} \Pi_{i=1}^{n} E\left(\exp i\lambda\xi_i\right) P(N_t = n)$$

y

$$\sum_{n=0}^{\infty} \Pi_{i=1}^{n} E\left(\exp i\lambda\xi_i\right) P(N_t = n) = \sum_{n=0}^{\infty} \Pi_{i=1}^{n} E\left(\exp i\lambda\xi_i\right) e^{-c}c^n/n!\,.$$

Pero por definición $E(\exp i\lambda\xi_i)$ es la función característica de ξ_i, pero como las variables aleatorias ξ son idénticamente distribuidas, tienen la misma función característica, que llamaremos $\varphi_\xi(\lambda)$, es decir

$$\varphi_\xi(\lambda) = E(\exp i\lambda\xi_i) \text{ para toda } i\,,$$

entonces si $\xi = \xi_1$

$$\sum_{n=0}^{\infty} \Pi_{i=1}^{n} E(\exp i\lambda\xi_i) e^{-c} c^n/n! = \sum_{n=0}^{\infty} \varphi_\xi(\lambda)^n e^{-c} c^n/n!$$

que se convierte, por simple agrupación

$$e^{-c}\left\{\sum_{n=0}^{\infty} (\varphi_\xi(\lambda)c)^n/n!\right\}$$

pero como conocemos el desarrollo de Taylor de la función exponencial tenemos que

$$e^{-c}\{\sum_{n=0}^{\infty}(\varphi_\xi(\lambda)c)^n/n!\} = e^{-c}\exp\{\varphi_\xi(\lambda)c = exp(-c(1-\varphi_\xi(\lambda)))\,,$$

de donde,

$$\varphi_X(\lambda) = \exp\{-c(1-\varphi_\xi(\lambda))\}\ ,$$

entonces

$$\psi(\lambda) = -c(\varphi_\xi(\lambda)+1)\,.$$

Pero tenemos que $-\varphi_\xi(\lambda)+1 = -E(\exp i\lambda\xi)+1 = -E(\exp i\lambda\xi+1) = \int_{\mathbb{R}} \exp(i\lambda x) + 1\gamma(dx)$ donde $\gamma(dx)$ es la distribución de ξ y c el parametro del poroceso Poisson.

$$\psi(\lambda) = c\int_{\mathbb{R}} \left(1 - e^{i<\lambda,x>}\right)\gamma(dx)\,.$$

Notemos que el todo proceso Poisson es caracterizado por (c,$\gamma(dx)$), lo que nos da una biyección entre las medidas finitas y los procesos de Poisson compuestos.

3.4 Formula de Lévy-Khinchine

Hemos visto que los procesos de Lévy tienen un exponente carcterístico, lo que veremos a continuacíon es que este exponente característico no puede ser de cualquier forma, tiene una forma definida que se explica en el siguiente teorema[2].

Teorema 3.10 (Formula de Lévy-Khinchin)**.** *El exponente caracteristico* $\psi\colon \mathbb{R} \to C$ *de cualquier proceso de Lévy se puede escribir como*

$$\psi(\lambda) = -ia\lambda + \frac{1}{2}Q^2\lambda^2 + \int_{\mathbb{R}-\{0\}} (1 - e^{ix\lambda} + ix\lambda I_{|x|\leq 1})\Pi(dx)\,,$$

donde $a \in \mathbb{R}$*,* $Q \geq 0$ *y* Π *es una medida sobre* $\mathbb{R}-\{0\}$ *tal que* $\int_{\mathbb{R}-\{0\}} \text{mín}(1, x^2)$ $\Pi(dx) < \infty$*.*

Inversamente, dada $a \in \mathbb{R}$*,* $Q \geq 0$ *y* Π *una medida sobre* $\mathbb{R} - \{0\}$ *tal que* $\int_{\mathbb{R}-\{0\}} \text{mín}(1, x^2)\Pi(dx) < \infty$*, para* $\lambda \in R$ *definimos*

$$\psi(\lambda) = -ia\lambda + \frac{1}{2}Q^2\lambda^2 + \int_{R-\{0\}} (1 - e^{ix\lambda} + ix\lambda I_{|x|\leq 1}\Pi(dx))\,.$$

Existe una única medida de probabilidad P *en* Ω *bajo la cual el proceso canónico* ξ *es un proceso de Lévy con exponente característico* $\psi(\lambda)$*.*

El coeficiente a es conocido como coeficiente lineal, Q es el coeficiente Gaussiano y Π es la medida de Lévy del proceso. El vector (a, Q, Π) caracteriza la ley del proceso.

Otra manera de escribir la formula es

$$\psi(\lambda) = -ia\lambda + \frac{1}{2}Q^2\lambda^2 + \int_{|x|\geq 1} (1 - e^{ix\lambda})\Pi(dx) + \int_{|x|\geq 1} (1 - e^{ix\lambda} + ix\lambda\Pi(dx))\,.$$

Es muy simple pasar de una a otra representacion, basta observar la indicadora. Esta manera de escribirla nos conviene pues hace mas claro que cualquier proceso de Lévy se puede ver como la suma de un término lineal, Q veces un proceso de Winer, un proceso Poisson Compuesto y una martingala M_t, es decir:

$$\xi_t = ct + QB_t + Y_t + M_t\,;$$

donde cada proceso es independiente de los demás. Un pequeño corolario de esto es que, como los saltos del proceso están en los últimos 2 términos de la fórmula de Lévy-Khinchin, la única familia de procesos de Lévy continuos

[2]La demostración se puede consultar en el libro *Procesos de Lévy* de Bertoi.

es la del movimientos Browniano multiplicados por una constante más un término de deriva.

$$ct + QB_t \,.$$

Ya calculamos los exponentes característicos de algunos procesos, entonces podemos dar explícitamente la tripleta de su formula de Lévy-Khinchine.

Para el movimiento Browniano tenemos que $\psi(\lambda) = \frac{1}{2}\lambda^2$ por lo tanto la tripleta es

$$(a = 0, Q = 1, \Pi = 0)\,.$$

Para el proceso Poisson tenemos que $\psi(\lambda) = c(1 - e^{i\lambda})$ por lo tanto la tripleta es

$$(a = 0, Q = 1, \Pi = c\delta_{\{1\}})$$

donde $\delta_{\{1\}})$ es la medida de Dirac con masa el el punto 1.

Para el Proceso Poisson compuesto tenemos que $\psi(\lambda) = \int_{\mathbb{R}}(1-e^{i\lambda x})\gamma(dx)$ por lo tanto la tripleta es

$$(a = 0, Q = 1, \Pi = \gamma(dx))\,.$$

Por último defininimos los procesos de Lévy en $\mathbb{R}^d$, como un proceso en $\mathbb{R}^d$ que cumple las 3 propiedades que le pedimos a un proceso de Lévy en $\mathbb{R}$. Definimos la función caracteristica como

$$\varphi_X(\lambda) = E(\exp i < \lambda, X >)\,,$$

y los procesos infinitamente divisibles en $\mathbb{R}^d$ tienen exponente caracterictico ψ definido como

$$\varphi_X(\lambda) = E(\exp i < \lambda, X >) = \exp(-t\psi(\lambda))\,.$$

Estos objetos nos servirán en el siguiente capítulo.

Capítulo 4

Convergencia en el espacio de Skorohod para procesos de Lévy

En este capítulo, concluimos la tesis probando un teorema que nos habla de convergencia, en el espacio de Skorohod, para procesos de Lévy. Para probar dicho teorema es indispensable usar el concepto de tensión, en particular el teorema de Prohorov, por lo que este capítulo junta los tres anteriores para darnos un bonito resultado de procesos de Lévy. Para tener convergencia de una sucesión $\{X^n\}$ de procesos de Lévy a otro proceso de Lévy en el espacio de Skorohod veremos que basta la convergencia débil de las variables aleatorias X_t^n a X_t para un tiempo fijo t, lo cual es un resultado sorprendente y útil.

Vamos a introducir algo de notación. Como ya lo hemos usado, si $\{Y_n\}_{n\geq 0}$ y Y son variables aleatorias que convergen débilmente escribiremos $Y_n \to Y$. Por otro lado si $\{X_n\}_{n\geq 0}$ y X son procesos estocásticos que convergen débilmente en el espacio de Skorohod escribiremos $X_n \Rightarrow X$.

El teorema que queremos probar es el siguiente.

Teorema 4.1. *Si X^n y X son procesos de Lévy, definidos en $(\omega, \Im, P)$, tales que $X_1^n \to X_1$, entonces se tiene la convergencia débil en el espacio de Skorohod*

$$X^n \Rightarrow X$$

Este resultado es muy importante.

ver [9]

Lo que debemos probar es convergencia finito dimensional y tensión para demostrar el teorema 1. Ahora probaremos un lema que nos sera útil, pero primero recordemos la desigualdad de Hölder: Si existen reales p y q mayores que uno, tales que $p^{-1}+q^{-1}=1$, funciónes $f_1 \in L_p$ y $f_2 \in L_q$ entonces $f_1 f_2 \in L_1$ y $E(f_1 f_2) \leq ||f_1||_p ||f_2||_q$.
Usando este teorema y el principio de inducción podemos generalizar la desigualdad en el siguiente sentido: si $f_1, f_2 ... f_n \in L_n$

$$E(f_1, f_2 ... f_n) \leq (E(f_1^n)^{\frac{1}{n}})(E(f_2^n)^{\frac{1}{n}})...(E(f_n^n)^{\frac{1}{n}})$$

Ahora si estamos en condiciones de entender el lema

Lema 4.2. *si $\xi_1 \dots \xi_n \dots \geq 0$ son variables aleatorias, y $S_n = \sum_1^n \xi_i$, se cumple que*

$$E[e^{-S_n}] \leq e^{-nc} + sup_{k \leq n} P\{\xi_k < c\}\,, \qquad c > 0\,.$$

Esto es fácil de ver usando primero la desigualdad de Hölder generalizada para tener

$$E[e^{-S_n}] = E[\Pi_1^n e^{-\xi_k}] \leq \Pi_1^n (E[e^{-n\xi_k}])^{\frac{1}{n}}\,,$$

y luego aplicanmos la desigualded de Chebyshev y separamos el caso de que que ξ_k sea mayor y que sea menor que c, para tener

$$\Pi_1^n (E[e^{-n\xi_k}])^{\frac{1}{n}} \leq e^{-nc} + sup_{k \leq n} P\{\xi_k < c\}\,,$$

de donde en efecto se cumple lo que queríamos.

Pensaremos que d es acotada, si no lo fuera se puede tomar la metrica equivalente $d(x,y) = \min(d'(x,y), 1)$.

Teorema 4.3 (Teorema de Aldous)**.** *Para $X^1, X^2, \dots, X^n, \dots$ elementos aleatorios de $D(\mathbb{R}^+, \mathbb{R})$, las afirmaciones 1,2 y 3 son equivalentes e implican 4.*

1. *Para toda sucesión $(\tau_n)_{t \geq n}$ de X^n-tiempos de paro opcionales y acotados y toda sucesión real h_n tal que $(h_n \to 0)$se cumple que*

   $$d(X^n_{\tau_n}, X^n_{\tau_n + h_n}) \to 0 \text{ si } n \to \infty$$

   en probabilidad.

2. $\lim_{\delta \to 0} \limsup_{n \to 0} \sup_{\tau \leq t} \sup_{h \in [0,\delta]} E(d(X^n_\tau, X^n_{\tau+h})) = 0$ *donde el primer supremo se toma sobre los X^n-tiempos de paro opcionales que son menores que t.*

3. $\lim_{h\to 0}\limsup_{n\to\infty}\sup_{\tau,\sigma} E(d(X_\tau^n, X_\sigma^n)) = 0$ *donde el supremo se toma sobre los tiempos de paro* σ, τ *tales que* $\sigma, \tau < t$ *y* $\sigma \leq \tau \leq \sigma + \tau$*.*

4. $\lim_{h\to 0}\limsup_{n\to\infty} E(\min(w_{ss}(X^n, t, h), 1)) = 0$*,* $t > 0$*.*

Demostración. 1) $\Longrightarrow$ 2) por el Teorema de Convergencia dominada en probabilidad.

2) $\Longrightarrow$ 1) de manera trivial.

Es obvio que 3) $\Longrightarrow$ 2).

2) $\Longrightarrow$ 3) Se muestra notando que si suponemos $0 \leq \tau - \sigma \leq \delta$, esto implica que $\tau \leq \sigma + \delta$, de donde

$$\sigma \leq \tau < \tau + \delta \leq \sigma + 2\delta\,,$$

o dicho de otro modo

$$[\tau, \tau + \delta] \subset [\sigma\, \sigma + 2\delta]\,.$$

Esta contención resultará muy importante en esta prueba, tambien nos será útil la contención

$$[\sigma, \tau + \delta] \subset [\sigma\, \sigma + 2\delta]\,.$$

Por la desigualdad del triángulo

$$d(X_\sigma^n, X_\tau^n) \leq d(X_\sigma^n, X_{\tau+h}^n) + d(X_\tau^n, X_{\tau+h}^n)\,,$$

integrando de ambos lados de cero a delta, con respecto a h, tenemos que

$$\delta d(X_\sigma^n, X_\tau^n) \leq \int_0^\delta (d(X_\sigma^n, X_{\tau+h}^n) + d(X_\tau^n, X_{\tau+h}^n))dh\,,$$

sin embargo, con la contención de conjuntos que mostramos y el hecho de que estamos integrando algo no negativo, tenemos que

$$\int_0^\delta d(X_\sigma^n, X_{\tau+h}^n)dh \leq \int_0^{2\delta} d(X_\sigma^n, X_{\sigma+h}^n)dh\,,$$

una forma de hacer intuitiva esta desigualdad es escribir

$$\int_0^{2\delta} d(X_\sigma^n, X_{\sigma+h}^n)dh = \int_0^{\tau} d(X_\sigma^n, X_{\sigma+h}^n)dh + \int_0^{\delta} d(X_\sigma, X_{\tau+h}^n)dh + \int_{\tau+\delta}^{2\delta} d(X_\sigma^n, X_{\tau+h}^n)dh\,,$$

entonces

$$\delta d(X_\sigma^n, X_\tau^n) \leq \int_0^\delta d(X_\tau^n, X_{\tau+h}^n)dh + \int_0^{2\delta} d(X_\sigma^n, X_{\sigma+h}^n))dh\,,$$

Sacando esperanza y usando Fubbini tenemos

$$\delta E[d(X^n_\sigma, X^n_\tau)] \leq \int_0^\delta E[d(X^n_\tau, X^n_{\tau+h})]dh + \int_0^{2\delta} E[d(X^n_\sigma, X^n_{\sigma+h}))]dh\,,$$

Pero podemos acotar superiormente a el lado izquierdo de la igualdad anterior tomando supremos de la siguiente forma

$$\begin{aligned} &\textstyle\int_0^\delta E[d(X^n_\tau, X^n_{\tau+h})]dh + \int_0^{2\delta} E[d(X^n_\sigma, X^n_{\sigma+h}))]dh \\ \leq &\textstyle\int_0^\delta \sup_{h\in[0,\delta]} E[d(X^n_\tau, X^n_{\tau+h})]dh + \int_0^{2\delta} \sup_{h\in[0,2\delta]} E[d(X_\sigma, X^n_{\sigma+h}))]dh \\ = &\,\delta \sup_{h\in[0,\delta]} E[d(X^n_\tau, X^n_{\tau+h})]dh + 2\delta \sup_{h\in[0,2\delta]} E[d(X^n_\sigma, X^n_{\sigma+h}))]dh\,, \end{aligned}$$

Lo que, cancelando las δ, implica que

$$E(X^n_\sigma, X^n_\tau) \leq \sup_{h\in[0,\delta]} E[d(X^n_\tau, X^n_{\tau+h})]dh + 2 \sup_{h\in[0,2\delta]} E[d(X^n_\sigma, X^n_{\sigma+h}))]dh\,,$$

Por último notemos que

$$\sup_{h\in[0,\delta]} E[d(X^n_\tau, X^n_{\tau+h})]dh \leq \sup_{h\in[0,2\delta]} E[d(X^n_\sigma, X^n_{\sigma+h})]dh$$

Pero esto nos permite concluir que

$$\sup_{\sigma,\tau} E(d(X^{\sigma, X^n_\tau))\leq 3\sup_\tau \sup_{h\in[0,2\delta]} E[d(X^n_\tau, X^n_{\tau+h})]}\,.$$

De lo anterior podemos concluir la implicación deseada, pues aplicando límite cuando δ tiende a cero y fijándonos en el límite superior, por hipótesis el lado derecho tiende a cero.

3) $\Longrightarrow$ 4).[1] Definimos los siguientes tiempos de paro:

$$\sigma^n_{k+1} = \inf\{s > \sigma^n_k : d(X^n_{\sigma^n_k}, X^n_s) > \epsilon\}\,;$$

con $\sigma^n_o = 0$. Es decir, esta familia de tiempos de paro crea una partición de $[0,t]$ tal que en cada segmento, ningún punto dista del primero más que ϵ, esto implica que ninguna pareja de puntos dista más de 2ϵ. Podemos asegurar la siguiente desigualdad (suponiendo que d es menor que 1 o tomando d'),

$$w_{ss}(X_n, t, h) \leq 2\epsilon + \sum_1^m I_{\{\sigma^n_{k+1} - \sigma^n_k < h, \sigma^n_k < t\}} + I_{\{\sigma^n_m < t\}}\,;$$

[1] (ver[6])

donde m es la cantidad de tiempos de paro sigma. La desigualdad es fácil de probar, si la partición generada por los tiempos de paro sigma es permitida, es decir $\sigma_{k+1}^n - \sigma_k^n < h$, entonces por construcción de la familia sigma, la oscilación es menor a 2ϵ, en el caso de que la partición no sea permitida, es decir que los tiempos de paro esten muy cerca, al menos una indicadora vale uno, y como la métrica está acotada por uno, la desigualdad se cumple.

Aplicando esperanza de ambos lados tenemos que

$$E[w_{ss}(X_n,t,h)] \leq 2\epsilon + \sum_{1}^{m} P\{\sigma_{k+1}^n - \sigma_k^n < h, \sigma_k^n < t\} + P\{\sigma_m^n < t\}\,,$$

pero notemos que la mayoria de los términos de la suma en la ecuación anterior se anulan cuando h tienda a cero. Esto es algo intuivo, pues cuando la h es muy pequeña se ocurre difícil que los tiempos de paro esten más cerca que h.

La manera formal de probar esto es ver que

$$P\{\sigma_{k+1}^n - \sigma_k^n < h, \sigma_k^n < t\} \leq \frac{1}{\epsilon}\sup E[d(X_\sigma, X_\tau)]\,;$$

donde el supremo es con los tiempos de paro menores que $t+h$ tales que $\sigma < \tau < \sigma + h$.[6] Esto implica que cuando h tiende a cero y n a infinito

$$P\{\sigma_{k+1}^n - \sigma_k^n < h, \sigma_k^n < t\} \to 0\,.$$

Lo anterior implica, tomando limites de ambos lados que

$$lim_{h\to 0}\limsup_{n\to\infty} E[w_{ss}(X_n,t,h)] \leq 2\epsilon + \limsup_{n\to\infty} P\{\sigma_m^n < t\}\,,$$

Por otro lado tenemos, usando el Lema 4.3 que

$$P\{\sigma_m^n < t\} \leq e^t E[e^{-\sigma_m^n}] : \sigma_m^n] \leq e^t\{e^{-mc} + \epsilon^{-1}\sup E[d(X_\sigma, X_\tau)]\}\,;$$

donde el supremo es con los tiempos de paro menores que $t+c$ tales que $\sigma < \tau < \sigma + c$. Entonces, por la afirmación 3 que tenemos como hipótesis

$$\limsup_{n\to\infty} P\{\sigma_m^n < t\} \to 0\,.$$

de donde finalmente concluimos que

$$\lim_{h\to 0}\limsup_{n\to\infty} E[w_{ss}(X_n,t,h)] \leq 2\epsilon\,,$$

lo que, por ser ϵ arbitrario, termina la prueba. $\square$

4.1 Convergencia finito dimensional

Para probar la convergencia finito dimensional usaremos el hecho de que la convergencia de funciónes características es necesaria y suficiente para asegurar la convergencia débil.

Por hipótesis tenemos que

$$X_1^n \to X_1 \text{ en distribución,}$$

por lo que

$$Lim_{n\to\infty}\varphi_{X_1^n}(\lambda) = \varphi_{X_1}(\lambda),$$

pero por ser procesos de Lévy, sabemos que

$$\varphi_{X_1^n}(\lambda) = E(exp(i\lambda X_1^n)) = exp(1\psi^n(\lambda))$$

y

$$\varphi_{X_1}(\lambda) = E(exp(i\lambda X_1)) = exp(1\psi(\lambda));$$

donde $\psi(\lambda)$ es el exponente característico.

Entonces tenemos que

$$Lim_{n\to\infty}exp(-\psi^n(\lambda)) = exp(-\psi(\lambda))$$

La prueba anterior sirve para todo t, es decir, prueba la convergencia unidimensional. No probaremos la convergencia finito dimensional en general, vamos a probarlo para un vector de tiempos (t_1, t_2),$t_1 < t_2$, para probarlo en general se usa la misma técnica.

Usaremos 2 argumentos, que son validos para todo proceso de Lévy $\{X_t\}_{t\geq 0}$:

- Se da la igualdad en ley$X_{t_2} - X_{t_1} = X_{t_2-t_1}$
- $X_{t_2} - X_{t_1}$ es independiente de X_{t_1}

Entonces notemos que para $\lambda = (\lambda_1, \lambda_2)$

$$E(exp(i < \lambda, (X_{t_1}^n, X_{t_2}^n - X_{t_1}^n) >)) = E(exp(i\lambda_1 X_{t_1}^n - i\lambda_2(X_{t_2}^n - X_{t_1}^n))) =$$
$$E((i\lambda_1 X_{t_1}^n)E((i\lambda_2 X_{t_2}^n - X_{t_1}^n) = E((i\lambda_1 X_{t_1}^n)E((i\lambda_2 X_{t_2-t_1}^n).$$

Donde la penúltima igualdad es por el argumento uno y la última por el 2. Por el capitulo de procesos de Lévy sabemos que

$$E((i\lambda_1 X_{t_1}^n)E((i\lambda_2 X_{t_2-t_1}^n) = exp(-t_1\psi^n(\lambda_1))exp(-(t_2-t_1)\psi^n(\lambda_2))$$

Y por la convergencia unidimencional que probamos se tiene que

$$exp(-t_1\psi^n(\lambda_1))exp(-(t_2-t_1)\psi^n(\lambda_2)) \to exp(-t_1\psi(\lambda_1))exp(-(t_2-t_1)\psi(\lambda_2))$$

Pero

$$exp(-t_1\psi(\lambda_1))exp(-(t_2-t_1)\psi(\lambda_2)) = E(i\lambda_1 X_{t_1})E(i\lambda_2 X_{t_2-t_1})$$

Ahora usamos que

$$E(i\lambda_2 X_{t_2-t_1}) = E(i\lambda_2(X_{t_2} - X_{t_1}))$$

y que

$$E(i\lambda_1 X_{t_1})E(i\lambda_2(X_{t_2} - X_{t_1})) = E(expi < \lambda, (X_{t_1}, X_{t_2} - X_{t_1}) >)$$

para determinar que

$$exp(-t_1\psi(\lambda_1))exp(-(t_2-t_1)\psi(\lambda_2)) = E(expi < \lambda, (X_{t_1}, X_{t_2} - X_{t_1}) >)$$

entonces

$$E(expi < \lambda, (X_{t_1}^n, X_{t_2-t_1}^n) >) \to E(expi < \lambda, (X_{t_1}, X_{t_2} - X_{t_1}) >).$$

De donde concluimos la siguiente convergencia debil

$$(X_{t_1}^n, X_{t_2}^n - X_{t_1}) \Rightarrow (X_{t_1}, X_{t_2} - X_{t_1})$$

Esto no es extactamente la convergencia bidimensional que buscabamos, para ello basta argumentar que la ecuación anterior y el hecho de que la función $f : \Re^2 \to \Re^2$, $f(x,y) = (x, x+y)$ es continua y el teorema 1.39 implican que

$$(X_{t_1}^n, X_{t_2}^n) \Rightarrow (X_{t_1}, X_{t_2})$$

Lo que nos asegura la convergencia bidimensional que buscabamos.

4.2 Tensión

Estamos en condiciones de probar que si X^n y X son procesos de Lévy tales que $Lim_{n\to\infty}X_1^n = X_1$, entonces se cumple que

$$d(X^n_{\tau_n}, X^n_{\tau_n+h_n}) \to 0 \text{ en probabilidad.}$$

y por el teorema de Aldous esto implica tensión, por lo que

$$X^n \Rightarrow X. \text{ en el sentido de la convergencia débil.}$$

Recordemos que X^n cumple la propiedad fuerte de Markov, por ser proceso de Lévy, entonces para cualquier sucesion de tiempos de paro τ_n y sucesión de constantes $h_n \to 0$, se cumple que si $X^n_{h_n} \to 0$ en distribución, entonces $X^n_{\tau_n+h_n} - X^n_{\tau_n} \to 0$, es decir

$$X^n_{h_n} \to 0 \text{ en distribución entonces } X^n_{\tau_n+h_n} - X^n_{\tau_n} \to 0 \text{ en distribución,}$$

pero además sabemos que convergencia en distribución a una constante implica convergencia en probabilidad a dicha constante, es decir

$$X^n_{\tau_n+h_n} - X^n_{\tau_n} \to 0 \text{ en distribución entonces } X^n_{\tau_n+h_n} - X^n_{\tau_n} \to 0 \text{ en probabilidad.}$$

Sabemos que convergencia en probabilidad a una constante implica convergencia en probabilidad del valor absoluto, es decir

$$X^n_{\tau_n+h_n} - X^n_{\tau_n} \to 0 \text{ en probabilidad entonces } |X^n_{\tau_n+h_n} - X^n_{\tau_n}| \to 0 \text{ en probabilidad.}$$

La convergencia del valor absoluto nos asegura la convergencia de la distancia, que es la hipótesis del teorema de Aldus. En resumen

$$X^n_{h_n} \to 0 \text{ en distribución entonces } |X^n_{\tau_n+h_n} - X^n_{\tau_n}| \to 0 \text{ en probabilidad.}$$

Entonces bastará probar que $X^n_{h_n} \to 0$ en distribución para terminar la prueba del teorema.

Para ver esto, recordemos que la convergencia de las funciónes características implica la convergencia en distribución, y que la función característica del cero es uno, por lo que queremos mostrar que

$$\varphi_{X^n_{h_n}}(\lambda) = E(exp(i\lambda X^n_{h_n})) = exp(-h_n\psi^n(\lambda)) \to 1,$$

pero nosotros sabemos que

$$\psi^n(\lambda)) \to \psi(\lambda)) < \infty,$$

pues $X_1^n \to X_1$. Además, por continuidad de la función exponencial tenemos que

$$lim_{n\to\infty} exp(-h_n\psi^n(\lambda)) = exp(-lim_{n\to\infty} h_n\psi^n(\lambda)) = e^0 = 1$$

lo que termina la prueba.

4.3 Aplicaciones

Finalmente veamos como se aplica el teorema 4.1.

Corolario 4.4. *Si ξ_n y ξ son procesos de Lévy con exponentes característicos φ_n y φ respectivamente y $\varphi_n \to \varphi$ entonces $\xi_n \Rightarrow \xi$*

La prueba es trivial, si convergen los exponentes característicos, tenemos convergencia débil para $t = 1$ y por el teorama de este capítulo tenemos convergencia de los procesos. Esto nos permite dar condiciones suficientes para que una sucesión de procesos converja a los procesos más importantes que hemos estudiado.

1. Movimiento Browniano. Una sucesión de procesos de Lévy X_n tales que $X_n = Q_n B_t + c_n t$, donde Q_n y c_n son constantes, converge a un movimiento browniano si $Q_n \to 1$ y $c_n \to 0$. La prueba es muy simple; calculemos el exponente característico de X_n

   $$E(e^{i(c_n+QB_1)\lambda}) = E(e^{iQB_1\lambda}e^{ic_n\lambda}) = e^{ic_n\lambda}E(e^{iQB_1\lambda}) = e^{ic_n\lambda}e^{\frac{1}{2}Q^2\lambda^2},$$

   de donde

   $$\varphi_n(\lambda) = ic_n\lambda + \tfrac{1}{2}Q^2\lambda^2$$

   de donde $X_n \to B_t$ si y sólo si $Q_n \to 1$ y $c_n \to 0$.

2. Proceso Poisson. Una sucesión de procesos Poisson de intensidad c_n converge a un proceso Poisson de intensidad c, si sólo si $c_n \to c$. La prueba es trivial.

3. Proceso Poisson compuesto. Para que una sucesión de procesos Poisson compuestos N_n^j converja a un proceso Poisson compuesto es suficiente que las funciónes de intensidad v_n converjan débilmente a v.

Ahora veamos 2 casos más interesantes.

- Sea B_t el movimiento Browniano. Sea X^n un proceso de Poisson de parámetro $\lambda_n = n$. Sea $\{\chi^n\}_{n\geq 0}$ una sucesión de procesos de Lévy tales que $\chi_1^n = \frac{1}{\sqrt{n}}(X_n - n)$, entonces

$$\chi^n \Rightarrow B$$

Demostración. Por ser procesos de Lévy y por el teorema de esta sección basta probar que $\chi_1^n \to B_1$, lo que es equivalente a probar que la distribución límite de $\chi_1^n = \frac{1}{\sqrt{n}}(X_n - n)$ es $N(0,1)$.

Para ver esto calcularemos la función característica de χ_1^n.

$$E(e^{i\lambda\chi_1^n}) = E(e^{i\lambda(\frac{1}{\sqrt{n}}(X_n - n))}) = e^{-i\lambda\sqrt{n}}E(e^{i\lambda(\frac{1}{\sqrt{n}}(X_n))}).$$

Pero $E(e^{i\lambda(\frac{1}{\sqrt{n}}(X_n))})$ se parece a la función característica de X_n que es una Poisson, así que la podemos calcular, $E(e^{i\lambda(\frac{1}{\sqrt{n}}(X_n))}) = e^{n(e^{\frac{-iu}{\sqrt{n}}}-1)}$, de donde

$$E(e^{i\lambda\chi_1^n}) = e^{-i\lambda\sqrt{n}}(e^{n(e^{\frac{-iu}{\sqrt{n}}}-1)}).$$

Ahora usemos la descomposición en serie de Taylor de $e^{\frac{-iu}{\sqrt{n}}}$

$$e^{\frac{-iu}{\sqrt{n}}} = 1 + \frac{iu}{\sqrt{n}} - \frac{u^2}{2n} - \frac{iu^3}{6n^{\frac{3}{2}}} + ... = \\ 1 + \frac{iu}{\sqrt{n}} - \frac{u^2}{2n} - \frac{h(u,n)}{\sqrt{n}};$$

donde h es un error que se mantiene acotado. Substituyendo esto tenemos que

$$E(e^{i\lambda\chi_1^n}) = e^{-i\lambda\sqrt{n}}(e^{n(e^{\frac{-iu}{\sqrt{n}}}-1)}) = \\ e^{-i\lambda\sqrt{n}}e^{n(1+\frac{iu}{\sqrt{n}}-\frac{u^2}{2n}-\frac{h(u,n)}{\sqrt{n}}-1)},$$

lo cual se simplifica en

$$E(e^{i\lambda\chi_1^n}) = e^{n(-\frac{u^2}{2n}-\frac{h(u,n)}{\sqrt{n}})},$$

de donde claramente

$$Lim_{n\to\infty}E(e^{i\lambda\chi_1^n}) = e^{-\frac{u^2}{2}}$$

que es la función característica de una Normal$(0,1)$, con lo que, en efecto, $\chi_1^n \to B_1$, lo que termina la prueba. $\square$

Definición 4.5. Diremos que una variable aleatoria Y es Poisson Compuesto si $Y = \xi_1$ donde $\{\xi_t\}_{t\geq 0}$ es un proceso Poisson compuesto.

- Toda medida de probabilidad infinitamente divisible se puede aproximar por una sucesión de medidas Poisson compuestas. [4] [10]

Demostración. Supongamos que X es una medida infinitamente divisible y x^n es el tal que $X = \sum_1^n x^n$. En el capitulo 3 justificamos que cuando $n \to \infty$, $x^n \to 0$ por lo que la función característica de x_n tiende a 1. Sea β la función característica de X y $\beta^{\frac{1}{n}}$ la de x^n, claramente $\beta = (\beta^{\frac{1}{n}})^n$. Por ser X infinitamente divisible tiene exponente característico y es $log(\beta)$, calculémoslo,

$$log(\beta) = nlog(\beta^{\frac{1}{n}}) \text{ para toda } n,$$

entonces la ecuación anterior tambien es valida para el límite cuando n tiende a infinito

$$log(\beta) = lim_{n\to\infty}nlog(\beta^{\frac{1}{n}}),$$

pero notemos que

$$lim_{n\to\infty}\frac{log((\beta^{\frac{1}{n}})(u)}{-(1-\beta^{\frac{1}{n}}(u))} = 1.$$

La ecuación anterior se prueba usando L'Hôpital. Notemos que la derivada de la parte de arriba del cociente es $\frac{1}{\beta^{\frac{1}{n}}(u)}\frac{d}{dx}\beta^{\frac{1}{n}}(u)$, mientras que la derivada de la parte de abajo es $\frac{d}{dx}\beta^{\frac{1}{n}}(u)$ por lo que el cociente de las derivadas es

$$\frac{1}{\beta^{\frac{1}{n}}(u)}\frac{d}{dx}\beta^{\frac{1}{n}}(u)/\frac{d}{dx}\beta^{\frac{1}{n}}(u) = \frac{1}{\beta^{\frac{1}{n}}(u)}$$

Lo cual claramente tiende a 1. Usando esto tenemos que

$$log(\beta) = -lim_{n\to\infty} n(1-\beta^{\frac{1}{n}}(u)),$$

pero como $\beta^{\frac{1}{n}}(u) = E(e^{iux_n}) = \int e^{iux} v_n(dx)$, donde v_n es la medida inducida por x_n, entonces podemos escribir

$$log(\beta) = -lim_{n\to\infty} n \int (1-e^{iux}) v_n(dx) = \\ -lim_{n\to\infty} \int (1-e^{iux})(nv_n)(dx)$$

Notemos además que $n\int(1-e^{iux})v_n(dx)$ es el exponente característico de un proceso Poisson compuesto, de donde en efecto la función característica de X es límite de funciónes características de variables aleatorias Poisson compuesto. □

Corolario 4.6. *Los procesos Poisson compuestos son un subconjunto denso de los procesos de Lévy.*

Demostración. Es concecuencia directa de que un proceso de Lévy parado al tiempo 1 es una variable aleatoria infinitamente divisible, del lema anterior y del teorema 4.1. □

4.4 2 maneras de ver a los procesos de Lévy

Trataremos al espacio de los procesos de Lévy como un espacio métrico. Lo compararemos con espacios que conocemos bien y de esta forma entenderemos mejor la estructura topológica del mismo.

Por un lado trabajaremos con la relación 1 a 1 entre los procesos de Lévy y las medidas de probabilidad infinitamente divisibles. El hecho de que las medidas de probabilidad infinitamente divisibles son subconjunto de las medidas de probabilidad nos permite usar la metrica de Prohorov sobre R para entender la topologia inducida por la convergencia débil sobre los procesos de Lévy.

Por otro lado estudiaremos al subespacio denso de los procesos de Lévy de los procesos Poisson compuestos. Aprovecharemos que su caracterización según la formula de Lévy-Kinchine es $(0, 0, \Pi)$ donde Π es una medida finita, (esto nos da una relación 1 a 1 entre los Procesos de Poisson compuestos y las medidas finitas).

4.4.1 Convergencia a tiempo 1

Sean A, B procesos de Lévy, definimos la métrica L_1

$$L_1(A, B) = \rho(A_1, B_1).$$

Hemos probado que la convergencia en tiempo 1 implica la convergencia de procesos, por lo que claramente

$$L_1(A_n, A) \to 0 \Leftrightarrow A_n \to A \text{ débilmente}$$

además de que L_1 es métrica porque ρ lo es. Esto prueba que el espacio de Lévy con la topología inducida por la convergencia débil es metrizable y homeomorfo al espacio metrico de las medidas infinitamente divisibles. Por ser homeomorfo al de las medidas infinitamente divisibles y por que éste está contenido en el de las medidas de probabilidad sabemos que es Housdorff.

4.4.2 Un denso: Los procesos Poisson compuestos

La función que manda a un proceso Poisson compuesto a su medida finita es un homeomorfismo, esto es una consecunecia directa de la caracterisación de Lévy-Kinchine. Por esta razón tenemos las siguiente propiedades topológicas de los procesos de Poisson:

Descripción topologica de los procesos Poisson compuestos

1. Es metrizable.
2. Es conexo.
3. Es separable.
4. Es primero y segundo numerable.
5. Es Housdorff.
6. No es localmente compacto.

Estas propiedades se pueden aprovechar para entender a los procesos de Lévy, pues el hecho de que los procesosde Poisson compuestos son un subconjunto denso implica que los procesos de Lévy son conexos, separables y primero y segundo numerables.

4.4.3 Descripción topologica de los procesos de Lévy

1. Es metrizable.
2. Es conexo.
3. Es separable.
4. Es primero y segundo numerable.
5. Es Housdorff.
6. No es localmente compacto.

4.5 Motivos de reflexión

Durante ésta tesis hemos podido entender una serie de relaciónes muy interesantes ente espacios relaciónados con la convergencia débil, recordemos estas relaciónes:

(a) El espacio de los procesos de Lévy y el de las medidas infinitamente divisibles son homeomorfos.

(b) Los procesos de Poisson compuestos, por medio de la fórmula de Lévy-Kinchine, son homeomorfas a las medidas finitas.

(c) Las medidas de Lévy y las medidas finitas sin mása en el cero son homeomorfas.

(d) las medidas finitas sin mása en el cero son densas en la medidas finitas y las medidas finitas son densas en las medidas de Lévy.

(e) Las medidas infinitamente divisibles son subconjunto de las medidas de probabilidad, que a su vez son subconjunto de las medidas finitas

(f) El espacio de los procesos de Poisson compuestos es subespacio denso de los procesos de Lévy.

Además de que estas relaciónes nos permiten asegurar que la topología de los espacios antes mencionados es muy parecida, la serie de contenciones y homeomorfismos nos invitan a conjeturar que todos esos espacios son homeomorfos. El análisis puntual de dichos espacios indica que la veracidad de esa conjetura es anti-intuitiva y casi escandaloso, por ejemplo, es claro que el movimiento Browniano no es, ni se parece a un proceso de Poisson compuesto, el espacio de los procesos de Lévy debería ser, entonces, muy diferente al de los procesos Poisson compuestos, sin embargo las anteriores relaciónes son difíciles de aceptar si hablamos de espacios topologicamente distintos. Ésta paradoja es motivo de reflexión.

Apéndice A

Topologia fuerte y topologia débil

Hemos estado trabajando con trayectorias de los reales a los reales, antes de profundizar en la métrica M_1, veremos como se trabaja en funciones de R a R^n y veremos algunos ejemplos.

Para pasar a R^n debemos notar que en la definición de ambas métricas estamos usando una métrica de las funciones de los reales a los reales, la del supremo, tenemos

$$d_{j_1T}(x,y) = \inf_{\lambda\in\Lambda} \max\{||x\circ\lambda - y||, ||e-\lambda||\},$$

y

$$d_{M1}(x,y) = \inf_{(u_1,r_1)\in\pi_1,(u_2,r_2)\in\pi_2} \max(||u_1-u_2||, ||r_1-r_2||).$$

En donde están las barras es donde debemos hacer el cambio, ahora la métrica del supremo no nos sirve y tenemos dos maneras de proceder, tomar la métrica del máximo

$$||x||_s = \max_{1\le i\le k} ||x^i||$$

donde $x = (x^1, \ldots, x^n)$. O bien, tomar la métrica del producto, que se llama así porque se deriva de pensar $R^n = R\times R\times\cdots\times R$

$$||x||_w = \sum_1^n ||x^i||,$$

donde $||x^i||$ es la métrica del supremo.

Tanto en J_1, como en M_1, usando estas nuevas métricas se guarda el comportamiento que hemos estado estudiando. Sin embargo, claramente la topología inducida por el uso de una u otra métrica es distinta, si utilizamos la del máximo se dice que estamos usando la topología fuerte y se denota J_s, o M_s según si estamos usando M_1 o J_1, (donde la s viene de la palabra en ingles strong), mientras que si usamos la métrica del producto inducimos la topología débil y la denotamos J_w o M_w respectivamente (por la palabre en ingles weak).

Estudiemos un ejemplo sencillo

Sea $X = (I_{[\frac{1}{2},1]}, I_{[0,\frac{1}{2})})$ y sea $X_n = (I_{[\frac{1}{2}-\frac{1}{n},1]}, I_{[0,\frac{1}{2}])} + \frac{1}{n})$, notemos que las funciones coordenadas convergen, la primera converge tanto en J_1 como en M_1 y la segunda converge incluso segun la métrica del supremo, así que usemos este ejemplo muy sencillo para sentirnos mas cómodos con las nueva métricas definidas.

En J lo que hay que hacer es construir la familia de homeomorfismos que muestran explícitamente la convergencia, sólo que esta vez hay que hacerlo entrada a entrada, la primera entrada ya la hemos estudiado antes, podemos usar los homeomorfismos que resultan de pegar la recta que va del $(0,0)$, al $(\frac{1}{2}-\frac{1}{n},\frac{1}{2})$, con la que va del $\left(\frac{1}{2}-\frac{1}{n},\frac{1}{2}\right)$ al (1,1), como ya hemos visto, estos homeomorfismos distan $\frac{1}{n}$ de la identidad, y hacen que $||x_n^1 \circ \lambda - x^1|| = 0$, en la otra entrada basta tomar la identidad pues $||x_n^2 - x^2|| = \frac{1}{n}$, de donde usando la topología fuerte

$$\begin{aligned} d_{j_s}(x, x_n) &= \inf_{\lambda\in\Lambda} \max\{||x_n \circ \lambda - x||, ||e - \lambda||\} \\ &\le \max\{\max(0, \frac{1}{n}), \max(\frac{1}{n}, 0)\} \\ &= \frac{1}{n}\,, \end{aligned}$$

de donde claramente

$$d_{j_s}(x, x_n) \to 0\,.$$

Ahora, si nos fijamos en la topologia débil tenemos

$$\begin{aligned} d_{j_w}(x, x_n) &= \inf_{\lambda\in\Lambda} \max\{||x_n \circ \lambda - x||, ||e - \lambda||\} \\ &\le \max\{(0 + \frac{1}{n}, \frac{1}{n} + 0) \\ &= \frac{1}{n}\,; \end{aligned}$$

de donde

$$d_{j_w}(x, x_n) \to 0\,.$$

Ahora veamos qué es lo que pasa con M. Notemos que en ambas entradas las funciones tienen una discontinuidad, sus gráficas completas constan de tres secciones, 2 lineas horizontales, es decir la gráfica de la función y una línea vertical que las une en la discontinuidad. las parametrizaciones que nos sirven son fáciles de ver, aquellas que recorren a la misma velocidad, cada una de las secciones, por ejemplo tomaremos las para metrizaciones que mandan el intervalo $\left[0, \frac{1}{3}\right)$ en la primera región, manda al intervalo $\left[\frac{1}{3}, \frac{2}{3}\right)$ en la segunda y el $\left[\frac{2}{3}, 1\right]$ en l tercera, en todas las X_n y también en X.

Fijamos n, estudiemos el comportamiento en la primera entrada, la diferencia en la coordenada temporal de la parametrización en la primera entrada, alcanza su máximo cuando el parámetro vale $\frac{1}{3}$, pues las respectivas parametrizaciones empiezan juntas, luego en el $\frac{1}{3}$ la de x se adelanta $\frac{1}{n}$, la diferencia en la coordenada temporal es cero, pues ambas parametrizaciones recorren el mismo camino en cada región, a la misma velocidad, de donde

$$||u^1_{X_n} - u^1_X|| = \frac{1}{n}, \quad ||r^1_{X_n} - r^1_X|| = 0\,.$$

En la otra entrada la coordenada espacial es la que se comporta igual en ambos casos, mientras que en la temporal, se alcanza el máximo en todos los puntos, siempre hay una distancia de $\frac{1}{n}$, por lo que

$$||u^1_{X_n} - u^1_X|| = 0, \quad ||r^1_{X_n} - r^1_X|| = \frac{1}{n}\,.$$

Ahora sí podemos calcular la distancia para mostrar explícitamente la convergencia también en M_1, primero con la topología fuerte

$$\begin{aligned} d_{Ms}(x, x_n) &= \inf_{(u_1,r_1)\in\pi_x,(u_2,r_2)\in\pi_{x_n}} \max(||u_1 - u_2||, ||r_1 - r_2||) \\ &\leq \max(||u_1 - u_2||, ||r_1 - r_2||) \\ &\leq \max(\max(\frac{1}{n}, 0), \max(0, \frac{1}{n})) \\ &= \frac{1}{n} \end{aligned}$$

Mientras que si usamos la métrica del producto tenemos

$$
\begin{aligned}
d_{Mw}(x, x_n) &= \inf_{(u_1,r_1)\in\pi_x,(u_2,r_2)\in\pi_{x_n}} \max(||u_1 - u_2||, ||r_1 - r_2||) \\
&\leq \max(||u_1 - u_2||, ||r_1 - r_2||) \\
&\leq \max(\frac{1}{n} + 0, 0 + \frac{1}{n}) \\
&= \frac{1}{n}
\end{aligned}
$$

De donde en ambos casos hay convergencia tal como queríamos.

Profundicemos en estas topologías, débil y fuerte, pero concentremonos en M_1.

Notemos que según la métrica que utilizamos la noción de segmento es diferente, vamos a definir el segmento producto:

si $a = (a^k, \ldots, a^k)$, $b = (b^1, \ldots, b^k)$, entonces el segmento producto, que denotaremos $[[a, b]]$

$$[[a, b]] = [a^1, b^1] \times \cdots \times [a^k, b^k] \, .$$

Donde $[a^i, b^i]$ es un segmento estándar, notemos que un segmento estándar en R^n es

$$[a, b] = \{\alpha a + (1 - \alpha)b : 0 \leq \alpha \leq 1\} \, .$$

Con estas nociones podemos definir las gráficas completas de una función en D, que nos ayudarán a entender mejor las topologías que induce M_1.

Sea x en D, para la topología fuerte definimos la gráfica delgada

$$\Gamma_x = \{(z, t) \in R^k \times [0, T] : z \in [x(t-), x(t)]\} \, .$$

Es decir, la gráfica delgada, consta de todos los puntos que están en la gráfica de la función x (recordamos que $x\colon R \to R^k$), más los segmentos que unen las discontinuidades de la gráfica.

Por otro lado definimos la gráfica gruesa

$$G_x = \{(z, t) \in R^k \times [0, T] : z \in [[x(t-), x(t)]]\} \, .$$

La diferencia es que ahora estamos tomando el hipercubo que resulta de hacer producto carteciano con todos los segmentos estándar, que se generan en cada entrada en las discontinuidades. Es decir, nuevamente tenemos la gráfica de la función, sólo que esta vez en lugar de unir las discontinuidades con segmentos estándar, las unimos con segmentos producto, otra manera de escribirlo es:

$$G_x = \{(z,t) \in R^k \times [0,T] : z^i \in [x(t-)^i, x(t)^i] \text{ para toda } i \in [1,...,k]\}.$$

Notemos que $[a,b] \subset [[a,b]]$, por lo que $\Gamma_x \subset G_x$.

Antes de proceder a las parametrizaciones debemos definir ordenes en las gráficas, lo cual ya no es tan intuitivo como cuando estabamos en R.

Nuestro orden es el siguiente, $(z_1,t_1) \leq (z_2,t_2)$ si

$$t_1 < t_2 \ \text{ o } \ t_1 = t_2 \text{ y } \ |x^i(t-) - z_1^i| \leq |x^i(t-) - z_2^i| \text{ para toda } i\,.$$

Es claro que éste no es un orden total en la gráfica gruesa, sin embargo eso no nos estorbará.

Otro concepto que usaremos es la proyección de la gráfica en R^k, es decir, quitar el parametro t y fijarnos en el resto de las coordenadas. A esto lo llamaremos rango y lo definimos de la siguiente manera:

El rango delgado de x es

$$\pi(\Gamma_x) = \{z \in R^k : (z,t) \in \Gamma_x \text{ para algun } t \in [0,T]\}.$$

El rango grueso de x es

$$\pi(G_x) = \Big\{z \in R^k : (z,t) \in G_x \text{ para algun } t \in [0,T]\Big\}.$$

Notemos que la pareja (z,t) está en Γ_x, para algún t, si sólo si z está en $\pi(\Gamma_x)$, análogamente la pareja (z,t) está en G_x, para algún t si sólo si z está en $\pi(G_x)$.

Ahora sí podemos definir parametrización y así conocer esta nueva caracterización de las topologías que induce M_1.

Definimos parametrización fuerte de x:

Una *parametrización fuerte* es una función continua y no decreciente que manda el intervalo $[0,1]$ al Γ_x, denotaremos (u,r) a la parametrización, donde $u = (u^1, \ldots, u^k) \in C([0,1], R^k)$ (donde $C([A],[B])$, es el conjunto de las funciones continuas que van del conjunto A al B) y $r = [0,1], [0,T]$.

Una parametrización débil de x es:

Una funcion continua y no decreciente p que manda el intervalo $[0,1]$ en el G_x, es importante notar que $p([0,1])$ está contenido en G_x, mas no necesariamente son iguales. (Recordemos que no decreciente es en el sentido del orden de la gráfica).

Ahora definimos $\Pi_s(x)$ al conjunto de todas las parametrizaciones fuertes de de x y ahora definimos $\Pi_w(x)$ al conjunto de todas las parametrizaciones débiles de x.

Notemos que en las funciones de los reales a los reales, el conjunto de las parameretrizaciones que habíamos definido anteriormente, coincide con el fuerte y el débil.

Ahora sí, usamos estos conjuntos para caracterizar las topologías M_1 fuerte y M_1 debil. La $||.||_k$ denota la metrica del supremo en R^k, $||x||_k = \sup_{0 \leq t \leq T}\{\max_{1 \leq i \leq k} |x^i(t)|\}$, entonces si $x_1, x_2 \in D$ definimos

$$d_s(x_1, x_2) = \inf_{(u_{1,2}, r_{1,2}) \in \Pi_s(x_{1,2})} \{\max\{||u_1 - u_2||_k, ||r_1 - r_2||\}\} ,$$

$$d_w(x_1, x_2) = \inf_{(u_{1,2}, r_{1,2}) \in \Pi_w(x_{1,2})} \{\max\{||u_1 - u_2||_k, ||r_1 - r_2||\}\} ;$$

donde la primera es la métrica sM_1 y la segunda no es métrica pero induce la topología wM_1, las topologías se inducen caracterizando la convergencia de la siguiente manera: En wM_1, $X_n \to X$ si sólo si $d_w(x, x_n) \to 0$.

Realmente lo que estamos haciendo es pegar las gráficas por las discontinuidades y así compararla, notemos que si estamos pensando en la topología fuerte diriamos: hay una discontinuidad, peguémosla con una recta y ya está. Mientras que si estamos en la débil diríamos: tenemos una discontinuidad, en efecto podemos pegarla con una recta, sin embargo no es la unica manera, vamos a aceptar todos los pegados que no se salgan de un hipercubo formado por los saltos en cada entrada, mientras sean crecientes en el sentido que definimos.

Trabajemos un sencillo ejemplo para aclimatarnos:

Ejemplo A.1. Trabajaremos en $D([0,2],R^2)$, analizaremos dos sucesiones muy parecidas, pero veremos que su comportamiento es completamente distinto,sea $X=(I_{[1,2]},I_{[1,2]})$, $X_n=(I_{[1+\frac{1}{n},2]},I_{[1+\frac{1}{n},2]})$ y $Y_n=(I_{[1-\frac{1}{n},2]},I_{[1+\frac{1}{n},2]})$, nosotros ya sabemos que sí convergen entrada a entrada, pues nos familiarizamos con la caracterización anterior de la convergencia débil y fuerte M_1, pero veamos como se trabaja aquí. Denotaremos al punto cuya entrada $x^1=a$, cuya entrada $x^2=b$ y cuyo componente temporal $t=t$, (t,a,b).

La gráfica completa delgada de X, consta del segmento de recta que une el punto $(0,0,0)$ con el $(1,0,0)$, el que une el $(1,0,0)$, con el $(1,1,1)$ y finalmente el que une el $(1,1,1)$ con el $(2,1,1)$.

La gráfica completa gruesa de X, es el segmento de recta que une el punto $(0,0,0)$ con el $(1,0,0)$, el rectángulo con vértices en $(1,0,0)$ y $(1,1,1)$ y el que segmento de recta une el $(1,1,1)$ con el $(2,1,1)$.

(a) La gráfica completa delgada de X_n, consta del segmento de recta que une el punto $(0,0,0)$ con el $\left(1-\frac{1}{n},0,0\right)$, el que une el $(1-\frac{1}{n},0,0)$, con el $(1-\frac{1}{n},1,1)$ y finalmente el que une el $(1,1,1)$ con el $(2,1,1)$.

La gráfica completa gruesa de X, es el segmento de recta que une el punto $(0,0,0)$ con el $((1-\frac{1}{n},0,0)$, el rectángulo con vértices en $(1-\frac{1}{n},0,0)$ y $(1-\frac{1}{n},1,1)$ y el que segmento de recta une el $(1-\frac{1}{n},1,1)$ con el $(2,1,1)$.

La parametrizacion de X_n consistirá en mandar el primer tercio del intervalo $[0,1]$ en el segmento de recta que une el punto $(0,0,0)$ con el $(1-\frac{1}{n},0,0)$, el segundo en el que une el $(1-\frac{1}{n},0,0)$, con el $(1-\frac{1}{n},1,1)$ y el último en el que une el $(1,1,1)$ con el $(2,1,1)$. Notemos que esta parametrizacion es permitida en la delgada y la gruesa.

La parametrizacion de X_n consistira en mandar el primer tercio del intervalo $[0,1]$ en el segmento de recta que une el punto $(0,0,0)$ con el $(1,0,0)$, el segundo en el que une el $(1,0,0)$, con el $(1,1,1)$ y el ultimo en el que une el $(1,1,1)$ con el $(2,1,1)$. Notemos que está parametrizacion también es permitida en la delgada y la gruesa.

Claramente la distancia de estas dos parametrizaciones es $\frac{1}{n}$, por lo que hay convergencia en ambas topologias.

(b) La gráfica completa delgada de Y_n, para n fija, es el segmento de recta que une el punto $(0,0,0)$ con el punto $(1-\frac{1}{n},0,0)$, luego la recta que une el $(1-\frac{1}{n},0,0)$ con el $(1-\frac{1}{n},1,0)$, después el que va del $(1-\frac{1}{n},1,0)$ al $(1+\frac{1}{n},1,0)$, después el segmento que une $(1+\frac{1}{n},1,0)$ con $(1+\frac{1}{n},1,1)$ y finalmente la que une $(1+\frac{1}{n},1,1)$ con $(2,1,1)$.

La gráfica completa gruesa de Y_n es igual a la delgada.

Gráficamente vemos que las gráficas delgadas de Y_n no se acercan a la gráfica delgada de X, en particular el punto $(1,\frac{1}{2},\frac{1}{2})$ que esta en la gráfica delgada de X, siempre esta lejos de la gráfica delgada de X_n, en el caso de la convergencia en la topología débil, vemos que la gráfica gruesa de X tiene una sección, que mas adelante definiremos como camino, al que las gráficas delgadas de X_n se estan acercando, esto nos permite sospechar que si habra convergencia en la topología débil.

Demos explícitamente una familia de parametrizaciones, que asegure la convergencia en la topología débil. Notemos que en las gráficas completas delgadas de las X_n hay 5 segmentos de rectas, así que dividimos el intervalo en 5. La parametrización de cada X_n manda el primer quinto en el segmento de recta que une el punto $(0,0,0)$ con el punto $(1-\frac{1}{n},0,0)$, el segundo quinto a la recta que une el $(1-\frac{1}{n},0,0)$ con el $(1-\frac{1}{n},1,0)$, el tercer quinto lo manda a segmento que va del punto $(1-\frac{1}{n},1,0)$ al $(1+\frac{1}{n},1,0)$, después el siguiente al segmento que une $(1+\frac{1}{n},1,0)$ con $(1+\frac{1}{n},1,1)$ y finalmente, el último quinto a la recta que une $(1+\frac{1}{n},1,1)$ con $(2,1,1)$.

La parametrizacion de X será de la siguiente manera: manda el primer quinto en el segmento que une el punto $(0,0,0)$ con el $(1,0,0)$, el segundo quinto se lo manda en el segmento que une el $(1,0,0)$ con el $(1,1,0)$ el tercer quinto se queda en ese punto, despues el siguiente quinto al segmento que une $(1,1,0)$ con $(1,1,1)$ y finalmente, el último quinto a la recta que une $(1,1,1)$ con $(2,1,1)$.

De esta manera podemos calcular $\max ||u_x, u_{x_n}||, ||r_x, r_{x_n}||$ para esta parametrización, notemos que en la primera sección, o el primer quinto se alcanza una distancia de $\frac{1}{n}$ en la parte temporal, en la segunda sección, de nuevo, sólo hay una distancia de $\frac{1}{n}$ en la parte temporal que se alcanza en todo punto de la región, en a tercera nuevamente solo hay una distancia de $\frac{1}{n}$ en la parte temporal que se alcanza en los extremos, nuevamente en la cuarta region hay una diferencia en la parte temporal, que se alcanza en todos los puntos y es $\frac{1}{n}$, y en la última entrada una vez más se alcanza una distancia de $\frac{1}{n}$ en la parte temporal, en el primer punto. La distancia en la parte espacial es cero siempre.

Entonces:

$$\begin{aligned} d_s(x, x_n) &= \inf_{(u_{1,2}, r_{1,2}) \in \Pi_s(x_n, x)} \{\text{máx}\{||u_1 - u_2||, ||r_1 - r_2||\}\} \\ &\leq \text{máx}\, ||u_x, u_{x_n}||, ||r_x, r_{x_n}|| \\ &= \text{máx}\{\frac{1}{n}, 0\} \\ &= \frac{1}{n} \end{aligned}$$

Notemos que esto no nos garantiza la convergencia fuerte, la parametrización de X que usamos no está en las permitidas para la convergencia fuerte.

De hecho, tal como indicaba la intuición gráfica no hay convergencia fuerte, basta notar que toda parametrización de X_n pasa por el punto $(1 - \frac{1}{n}, 1, 0)$, como este punto dista 1 de la grafica delgada de X, cualquier pareja de parametrizaciones de X y X_n, en t tal que $(u_n(t), r_(t)) = (1 - \frac{1}{n}, 1, 0)$ cumplen que

$$|u(t), u_n(t)| \geq 1\,,$$

por lo que en efecto no hay convergencia en la topología fuerte.

Notemos que todas las parametrizaciones que están en $\Pi_s(x)$ están en $\Pi_w(x)$ pues $[a, b] \subset [[a, b]]$, entonces $d_s(x_1, x_2) \leq d_w(x_1, x_2)$ para toda pareja $x_1, x_2 \in D$, esto nos explica el nombre, fuerte y débil, pues convergencia en el fuerte implica en el débil más alrevés no. Esto lo veremos en el siguiente ejemplo.

Ejemplo A.2. En $D([0, 2], R^2)$ de nuevo, sea $X = (I_{[1,2]}, I_{[1,2]})$, y sea $X_n = (2I_{[1-\frac{1}{n},2]}, I_{[1,2]} + I_{1-\frac{1}{n},1})$.

Primero describamos las gráficas completas delgadas: La de X_n, consta de la recta que va del $(0, 0, 0)$ al $(1 - \frac{1}{n}, 0, 0)$, la recta que va de $(1 - \frac{1}{n}, 0, 0)$ a $(1 - \frac{1}{n}, 2, 1)$, la que va de $(1 - \frac{1}{n}, 2, 1)$ a $(1, 2, 1)$, la que va de $(1, 2, 1)$ a $(1, 2, 2)$ y finalmente la que va de $(1, 2, 2)$ a $(2, 2, 2)$.

La de X es el segmento que une el $(0,0,0)$ con el $(1,0,0)$, el que une en $(1,0,0)$ con el $(1,2,2)$ y el que une este último con el $(2,2,2)$.

Fijémonos en los rangos, notemos que la recta que une el punto $(0,0)$ con el $(2,1)$ unido con la que ve de este último al $(2,2)$ conforma el rango delgado de todas las X_n, mientras que el rango delgado de X es la recta que une el $(0,0)$ con el $(2,2)$.

La parte espacial de toda parametrización de cualquier X_n tiene que pasar por el punto $(2,1)$, y la distancia de este punto a la identidad es uno (segun la métrica que estamos usando), por lo que tenemos:

$$d_s(x, x_n) = \inf_{(u_{1,2}, r_{1,2}) \in \Pi_s(x_n, x)} \{\max\{||u_1 - u_2||, ||r_1 - r_2||\}\} \geq 1 > 0\,;$$

de donde en efecto no hay convergencia fuerte.

Ahora describamos la gráfica gruesa: La de X es el segmento que une el $(0,0,0)$ con el $(1,0,0)$, el rectángulo con vértices en $(1,0,0)$ y $(1,2,2)$, y el segmento que va del $(1,2,2)$ al $(2,2,2)$.

La de X_n, consta de la recta que va del $(0,0,0)$ al $(1-\frac{1}{n},0,0)$, el rectángulo con vértices en $(1-\frac{1}{n},0,0)$ y $(1-\frac{1}{n},2,1)$, la recta que va de $(1-\frac{1}{n},2,1)$ a $(1,2,1)$, la que va de $(1,2,1)$ a $(1,2,2)$ y finalmente la que va de $(1,2,2)$ a $(2,2,2)$.

Antes de empezar los cálculos con wM_1 convenzámonos de que puede haber convergencia en ese sentido.

De nuevo fijémonos en el rango, esta vez en el rango grueso, el rango grueso de X es todo el cuadrado $[0,2] \times [0,2]$, mientras que el de X_n es el rectangulo $[0{,}2] \times [0,1]$ y la recta que va del $(2,1)$ al $(2,2)$.

Es claro que el rango grueso de X, contiene a todos los rangos gruesos de las X_n. Esto es una buena noticia, además notamos que la coordenada temporal se va acercando, así que es intuitivo que se dé la convergencia.

Construyamos ahora una parametrización que nos asegure la convergencia: Tomamos las gráficas delgadas de X_n, parametrizarlas, y luego construir una parametrización de la gráfica gruesa de X, que haga explícita la convergencia.

La parametrización manda el intervalo $[0,1]$, en Γ_{x_n} de la siguiente manera, al primer quinto lo manda en la recta que va del $(0,0,0)$ al $(1-\frac{1}{n},0,0)$, el segundo en la recta que va de $(1-\frac{1}{n},0,0)$ a $(1-\frac{1}{n},2,1)$, el tercero en la que va de $(1-\frac{1}{n},2,1)$ a $(1,2,1)$, el cuarto la que va de

$(1,2,1)$ a $(1,2,2)$ y finalmente el último quinto en la que va de $(1,2,2)$ a $(2,2,2)$.

Ahora construiremos una parametrización de X, que "cache" las parametrización de las X_n, aprovechando que la gráfica gruesa de X, nos da muchas opciones. Procedemos de la siguiente manera manda el primer quinto en el segmento que une el punto $(0,0,0)$ con el $(1,0,0)$, el segundo quinto en el segmento de recta que une el $(1,0,0)$ con el $(1,2,1)$, el tercer quinto se queda en ese punto, después el siguiente quinto al segmento que une $(1,2,1)$ con $(1,2,2)$ y finalmente, el último quinto a la recta que une $(1,2,2)$ con $(2,2,2)$.

Notemos que esta es una parametrización permitida, pues es no decreciente, con respecto al orden de la gráfica y está contenida en la gráfica gruesa.

Así para una n fija, tenemos que en todos los quintos se alcanza una distancia máxima de $\frac{1}{n}$ en la coordenada temporal, y nunca hay una distancia positiva en la coordenada espacial.

Entonces

$$\begin{aligned} d_w(x,x_n) &= \inf_{(u_{1,2},r_{1,2})\in\Pi_s(x_n,x)} \{\max\{||u_1-u_2||,||r_1-r_2||\}\} \\ &\leq \max ||u_x,u_{x_n}||,||r_x,r_{x_n}|| \\ &= \max\{\frac{1}{n},0\} \\ &= \frac{1}{n}\,; \end{aligned}$$

lo que hace explícita la convergencia.

Ahora que hemos visto bastante de las topologías fuerte y débil, estamos listos para el siguiente ejemplo, el cual nos ilustrará que, contrario a la intuición, d_w no es métrica.

Ejemplo A.3. Retomemos el ejemplo anterior, estamos en $D([0,2],R^2)$, sea $X=(I_{[1,2]},I_{[1,2]}$, y sea $X_n=(2I_{[1-\frac{1}{n},2]},I_{[1,2]}+I_{[1-\frac{1}{n},1]})$. Ya concluimos que $d_w(X,X_n)\to 0$, ahora definamos $Y_n=(I_{[1,2]}+I_{[1-\frac{1}{n},1]},2I_{[1-\frac{1}{n},2]})$.

Notemos que es igual a X_n sólo que pasamos lo de una coordenada a la otra, así que si hacemos exactamente los mismos pasos que en el ejercicio anterior, pero cambiando las coordenadas, concluiriamos que $d_w(X,Y_n)\to 0$.

Lo que esto significa es que para toda $\epsilon > 0$ existe un elemento de X_n y uno de Y_n tales que $d_w(Y_0, X) < \epsilon$ y $d_w(X_0, X) < \epsilon$, entonces para que se cumpla la desigualdad del triángulo tendría que pasar que $d_w(X_0, Y_0) < 2\epsilon$. Tomemos $\epsilon < \frac{1}{2}$.

Notemos que el rango grueso de Y_0 consiste en el rectángulo $([0,1] \times [0,2])$ y la recta que va del $(2,1)$ al $(2,2)$, mientras que el rango grueso de X_0 consta del $([0,2] \times [0,1])$ y la recta que va del $(1,2)$ al $(2,2)$.

Esto nos hace pensar que su distancia no es menor que 2ϵ, y en efecto no lo es, notemos que toda parametrización de Y_0 debe pasar por $(1, 2, 1 - \frac{1}{k})$, y ya conocemos la gráfica gruesa de X_0, lo más que se acerca a este punto en las coordenadas espaciales, segun esta métrica es 1, por lo que la distancia $d_w(X_0, Y_0) > 1$.

Entonces

$$d_w(X, X_0) + d_w(X, Y_0) < 2\epsilon < 1 \leq d_w(X_n, Y_0)$$

de donde

$$d_w(X, X_0) + d_w(X, Y_0) \leq d_w(X_n, Y_0)$$

lo que prueba que no se cumple la desigualdad del triángulo.

El ejemplo anterior no sólo nos demostró que d_w no es una métrica, sino que también nos sembró la duda de si d_s es métrica. La respuesta es sí, pero la prueba tiene que esperar algunos lemas y definiciones.

Definición A.4. Sea $A \subset \Gamma_w(x)$ un conjunto de puntos ordenados (segun el orden de la gráfica), tales que el primero es el $(x(0), 0)$ y el último el $(x(T), T)$ la distancia consistente con el orden de una gráfica es:

$$d^{\cdot}(A, \Gamma_x) =$$

$$\sup_{(z_i,t_i)\in A} \sup_{(z,t)\in[(z_i,t_i)(z_{i+1},t_{i+1})]} \text{máx}(||(z,t)-(z_i,t_i)||, ||(z,t)-(z_{i+1},t_{i+1})||)\,.$$

Tanto el siguiente lema como la anterior definición parecen bastante inútiles, la verdad es que no lo son, con ellos construiremos la prueba de que d_s es métrica, una vez hecha esta defensa seguimos con el lema.

Lema A.5. *Para cualquier $x \in D$ y cualquier $\epsilon > 0$, existe un conjunto ordenado $A \subset \Gamma_x$, tal que $d^{\cdot}(A, \Gamma_x) < \epsilon$.*

La prueba es muy simple, para formar a A tomemos el primer punto de la gráfica, luego caminamos ϵ, en el sentido de la distancia usual de R^k, sobre la gráfica (como estamos hablando de gráficas delgadas sólo hay un camino posible) y luego caminamos otro ϵ y tomamos el segundo punto, así seguimos hasta llegar al final de la gráfica, ponemos el último punto como el punto $(x(T), T)$, notemos que como la gráfica es continua, sólo pusimos una cantidad finita de puntos, y es claro que $d'(A, \Gamma_x) < \epsilon$ por lo que se concluye lo deseado.

Ahora el último lema antes de pasar a la prueba que nos incumbe.

Lema A.6. *Para cualquier pareja* $x_1, x_2 \in D$ *y cualquier parametrización* $(u_1, r_1) \in \Pi_s(x_1)$*, existe una parametrización* $(u_2, r_2) \in \Pi_s(x_2)$ *tal que*

$$\text{máx}(|u_1, u_2|, |r_1, r_2|) < d_s(x_1, x_2) + \epsilon\,.$$

Demostración. Podemos tomar una pareja de parametrizaciones tales que el máximo de la distancia de sus entradas sea menor que cualquier constante, por el ínfimo en la definición de la metrica M_1, eso haremos, tomamos $(u'_1, r'_1) \in \Pi_s(x_1)$ y $(u'_2, r'_2) \in \Pi_s(x_2)$ tales que

$$\text{máx}(|u'_1, u'_2|, |r'_1, r'_2|) < d_s(x_1, x_2) + \frac{\epsilon}{3}\,.$$

Ahora definimos el conjunto finito A, un conjunto de puntos tales que $d'(A, \Gamma_x) < \frac{\epsilon}{3}$, como es finito podemos encontrar otro conjunto con la misma cantidad de puntos, ordenado del mas chico al más grande, contenido en $[0, 1]$, al que llamaremos S'_1, tal que $s'_i \in S'_1$ cumple que $(u'_1(s'_i), r'_1(s'_i))$ es el i-ésimo elemento de A, definimos S_1 de la misma manera que S'_1 con la diferencia de que $(u_1(s_i), r_1(s_i))$ es el i-ésimo elemento de A. Sea λ un homeomorfismo del $[0, 1]$ en si mismo, que manda los elementos de el i-ésimo elemento de S_1 en el $i-$ésimo de S'_1. Sea $(u_2, r_2) = (u'_2 \circ \lambda, r'_2 \circ \lambda)$. Notemos entonces que

$$\text{máx}(|u_1 - u_2|, |r_1 - r_2|) = \text{máx}(|u_1 - u'_1 \circ \lambda - u_2 + u'_1 \circ \lambda|, |r_1 - r_1^- \circ \lambda - r_2 + r'_1 \circ \lambda|) \leq \text{máx}(|u_1 - u'_1$$

Analizemos cada sumando del último renglón por separado;

$$\text{máx}(|u_2 + u'_1 \circ \lambda|, |r_2 - r'_1 \circ \lambda|) = \text{máx}(|u'_2 \circ \lambda + u'_1 \circ \lambda|, |r'_2 \circ \lambda - r'_1 \circ \lambda|) < d_s(x_1, x_2) + \frac{\epsilon}{3},$$

todo por construccion y por ser λ un homeomorfismo.

Por otro lado, para analizar el otro sumando usaremos que $u_1(s) - u'_1 \circ \lambda(s)$ para todo $s \in S_1$ y que $d'(A, \Gamma_x) < \frac{\epsilon}{3}$.

Sea s en $[0,T]$, sea s_i el más cercano elemento de S_1, menor que s. Entonces

$$|u_1(s) - u_1^{,} \circ \lambda(s)| = |u_1(s) - u_1(s_i) + u_1^{,} \circ \lambda(s_i) + u_1(s_i) - u_1^{,} \circ \lambda(s_i) - u_1^{,} \circ \lambda(s)|$$

$$\leq |u_1(s) - u_1(s_i)| + |u_1^{,} \circ \lambda(s_i) - u_1(s_i)| + |u_1^{,} \circ \lambda(s_i) - u_1^{,} \circ \lambda(s)|$$

$$\leq \frac{1}{3}\epsilon + 0 + \frac{1}{3}\epsilon \, ,$$

análogamente

$$|r_1(s) - r_1^{-} \circ \lambda(s)| \leq \epsilon \, ,$$

de donde

$$\text{máx}(|u_1 - u_1^{,} \circ \lambda|, |r_1 - r_1^{-} \circ \lambda|) \leq \epsilon \, ;$$

lo que nos permite concluir que $\text{máx}(|u_1 - u_2|, |r_1 - r_2|) < d_s(x_1, x_2) + \epsilon$. □

Ya tenemos todo lo necesario para probar que d_s es métrica.

Demostración. Claramente cumple que es positiva, reflexiva y simétrica, por ser $|.|$ métrica, sólo probaremos la desigualdad del triángulo.

Sea $(u_3, r_3) \in \Pi_s(x_3)$, tomamos $(u_1, r_1) \in \Pi_s(x_1)$ y $(u_2, r_2) \in \Pi_s(x_2)$ tales que

$$\text{máx}\, ||u_3 - u_1|| ||r_3 - r_1|| < d_s(x_3, x_1) + \frac{\epsilon}{2}$$

$$\text{máx}\, ||u_3 - u_2|| ||r_3 - r_2|| < d_s(x_3, x_2) + \frac{\epsilon}{2} \, ,$$

los cuales existen por el lema anterior.

Entonces

$$\begin{aligned} d_s(x_1, x_2) &\leq \text{máx}\, ||u_1 - u_2|| ||r_1 - r_2|| \\ &\leq \text{máx}\, ||u_3 - u_2|| ||r_3 - r_2|| + \text{máx}\, ||u_3 - u_1|| ||r_3 - r_1|| \\ &\leq d_s(x_2, x_1) + d_s(x_3, x_1) + \epsilon \, . \end{aligned}$$

Como fue para ϵ arbitraria haciendo esta tan chica como queramos obtenemos que:

$$d_s(x_1, x_2) \leq d_s(x_2, x_1) + d_s(x_3, x_1) \, .$$

□

Corolario A.7. *Como en el caso de las funciones de $\mathbb{R}$ a $\mathbb{R}$, la topología fuerte y la débil son la misma, tenemos que $M_1 = M_s$ en $\mathbb{R}$, es en efecto métrica.*

Extenderemos el concepto de osilacion que expusimos para la topologia m_1 en el caso real, de esta forma se extienden los resultados sobre tensión que también expusimos.

Las siguientes oscilaciones dependen de si estamos hablando de la topología fuerte o la débil, recordemos que [], denota el segmento normal y [[]] el segmento grueso de la topología débil.

Para la topologia fuerte

$$w_s(x,t,\delta) = \sup_{\text{máx}(0,t-\delta)\geq t_1\geq t_2\geq t_3\geq \text{mín}(T,t+\delta)} \{||x(t_2) - [x(t_1), x(t_3)]||\}\,,$$

para la débil

$$w_w(x,t,\delta) = \sup_{\text{máx}(0,t-\delta)\geq t_1\geq t_2\geq t_3\geq \text{mín}(T,t+\delta)} \{||x(t_2) - [[x(t_1), x(t_3)]]||\}\,.$$

Lo que hace esta función es fijarse en que tan lejos esta un punto de la gráfica de los segmentos que unen un punto anterior con uno posterior, en el orden de la gráfica. La noción de segmento es lo que cambia de una a otra.

Definimos

$$w_s(x,\delta) = \sup_{0\leq t\leq T} w_s(x,t,\delta)\,,$$

y

$$w_w(x,\delta) = sup_{0\leq t\leq T} w_w(x,t,\delta)\,.$$

Finalmente definimos

$$O_s = \text{máx}(w_s(x,\delta), v(x,0,\delta), v(x,T,\delta))\,,$$

y

$$O_w = \text{máx}(w_w(x,\delta), v(x,0,\delta), v(x,T,\delta))\,.$$

Las O nos hablan del comportamiento global de la función x y nos permiten enunciar el siguiente teorema.

Teorema A.8. *Una sucesión de medidas de probabilidad P_n sobre D, con la topologia SM_1 es tensa si y solo si se cumplen las siguientes 2 propiedades:*

(a) *para todo* $\epsilon > 0$, *existe* k *tal que*

$$P_n(\{x \in D : ||x|| > c\}) < \epsilon \text{ para toda } n\,.$$

(b) *Para todo* $\epsilon > 0$, $k > 0$ *existe* δ *tal que*

$$P_n(\{x \in D : O_s(X_n, \delta) \geq \epsilon) \leq k \text{ para toda } n\,.$$

En el caso de la topologia débil tenemos el siguiente teorema:

Teorema A.9. *Una sucesion de medidas de probabilidad* P_n *sobre* D, *con la topología* WM_1 *es tensa si y solo si se cumplen las siguientes 2 propiedades:*

(a) *para todo* $\epsilon > 0$, *existe* k *tal que*

$$P_n(\{x \in D : ||x|| > c\}) < \epsilon \text{ para toda } n\,.$$

(b) *Para todo* $\epsilon > 0$, $k > 0$ *existe* δ *tal que*

$$P_n(\{x \in D : O_w(X_n, \delta) \geq \epsilon) \leq k \text{ para toda } n\,.$$

Apéndice B

Tensión en el espacio de las funciones continuas

El siguiente teorema es para el subconjunto $C \subset D$, el de las funciones continuas.

Teorema B.1 (Arzela-Ascoli)**.** *Un subconjunto A de C es relativamente compacto si y sólo si*

$$\sup_{x \in A} x(0) < \infty \quad y \quad \lim_{\delta \to 0} \sup_{x \in A} v(x, \delta) = 0 \,.$$

Donde v es el modulo de continuidad que definimos como

$$v(x, \delta) = \sup\{|x(t_1) - x(t_2)| : 0 \geq t_1 < t_2 \geq 1, |t_1 - t_2| < \delta\} \,.$$

Este teorema nos permite entender el siguiente teorema, que nos permite entender la tensión de las variables aleatorias, al menos C.

Teorema B.2. *Una sucesion de variables aleatorias X_n en C, es tensa si y solo si se cumplen las siguientes 2 propiedades:*

(a) Para toda $\epsilon > 0$ existe una constante c tal que

$$P(|X_n(0)| > c) < \epsilon \text{ para toda } n \,.$$

(b) Para todo $\epsilon > 0$, $k > 0$ existen δ y $N_{(\epsilon,k)}$ tales que

$$P(v(X_n, \delta) \geq \epsilon) \leq k \quad \text{para toda} \quad n > N_{(\epsilon,k)} \,.$$

Esto nos enseña a saber si un conjunto de variables aleatorias en C es tenso.

Bibliografía

[1] J. Bertoin: Lévy processes, Cambridge University, Cambridge, 1996.

[2] P. Billingsley: Convergence of probability mesures, J. Wiley, New York, 1968.

[3] S. Ethier, T. Kurtz: Markov processes: Chartacterization and convergence, J. Wiley, New York, 1986.

[4] B. Fristedt, L. Gray: A modern approach to probability theory, Birkhauser, Boston, 1996.

[5] J. Jacod, P. Protter: Essentials of Probability, Springer-Verlag, New York, 2000

[6] O. Kallenberg: Foundations of Modern Probabiity, Springer, New York, 2002.

[7] D. Revuz and M. Yor: Continuous Martingales and Brownian Motion, Springer, New York, 1994.

[8] Tesis para obtener el grado de matemático de G. Uribe, F. Ciencias. UNAM.

[9] W. Whitt: Stochastic-Prosess Limits, Springer, 2002.

[10] K. Sato: Lévy Processes and Infinitely Divisible Distributions, Cambridge University, Cambridge, 1999.

[11] C. Prieto: Topología básica, FCE, México, 2003.