



UNIVERSIDAD
DE GRANADA



TRABAJO FIN DE MÁSTER

MÁSTER EN MATEMÁTICAS

Polinomios Ortogonales

Autor

Juan Antonio Villegas Recio

Tutoras:

Lidia Fernández Rodríguez

Departamento de Matemática Aplicada

Antonia María Delgado Amaro

Departamento de Matemática Aplicada

21 de junio de 2023



AGRADECIMIENTOS

Ya veremos a quién hay que agradecerle qué.

DECLARACIÓN DE AUTORÍA

Yo, D. Juan Antonio Villegas Recio, autor de este TFM,

Declaro explícitamente que el trabajo aquí presentado como Trabajo de Fin de Máster de mis estudios en el Máster Interuniversitario en Matemáticas, correspondiente al curso académico 2022-2023, es original. Es decir, no se han utilizado para el desarrollo del trabajo fuentes sin haberlas citado debidamente.

Granada, a 21 de junio de 2023.

A handwritten signature in black ink, appearing to read 'Juan Antonio', is written over a horizontal line.

Juan Antonio Villegas Recio

La aproximación en matemáticas siempre ha sido uno de las principales alternativas ante la dificultad de encontrar soluciones exactas a algunos problemas. En este marco se sitúan los polinomios ortogonales, útiles para, entre otros ámbitos, aproximar funciones o calcular integrales y transformadas.

Por otro lado, el interés y la necesidad de predecir sucesos futuros ha provocado que se desarrollen múltiples modelos probabilísticos, entre los que hoy destacamos las cadenas de Markov. En particular, las cadenas y procesos de Markov son modelos de estados, es decir, tratan de predecir la situación en la que se encontrará en un futuro un sistema a partir de muy poca información.

Podría parecer que hasta el momento se ha hablado de dos áreas muy distintas de las matemáticas, pues tanto su naturaleza como sus objetivos son notablemente diferentes. Sin embargo, existe una relación muy estrecha entre los polinomios ortogonales y las cadenas de Markov. Esta relación reside en los procesos de nacimiento y muerte, que son un caso particular de cadenas de Markov sobre las cuales estudiaron los matemáticos Samuel Karlin (1924–2007) y James McGregor (1926–1991).

En este trabajo estudiamos en principio de forma independiente los polinomios ortogonales y las cadenas de Markov para finalmente exponer la relación entre ambas. Así, la memoria de este trabajo se ha estructurado en cuatro capítulos:

1. **Introducción a los Polinomios Ortogonales:** Se explican los conceptos y resultados más elementales de la teoría de polinomios ortogonales, se establece la notación que se usará a lo largo del trabajo y se exponen algunos ejemplos clarificadores.
2. **Polinomios Ortogonales Clásicos:** Se presentan los polinomios ortogonales conocidos como clásicos, que son familias particulares con varias propiedades exclusivas que las caracterizan.
3. **Cadenas de Markov:** Ya introducida la teoría de polinomios ortogonales, se estudian desde un punto de vista puramente probabilístico las cadenas de Markov y su comportamiento, exponiendo diversos ejemplos.

4. **Polinomios Ortogonales y Procesos de Nacimiento y Muerte:** Una vez explicadas las dos disciplinas preliminares, se definen los procesos de nacimiento y muerte y se deduce la relación entre estos modelos probabilísticos y los polinomios ortogonales.

Además, a lo largo del documento se muestran ejemplos para los cuales se utilizaron herramientas de software matemático con el fin de ilustrar de manera práctica los conceptos teóricos. Mediante el uso de **SageMath** y de **R**, se llevaron a cabo cálculos simbólicos y numéricos, visualizaciones, medidas temporales y comparaciones que acercan al lector a casos concretos de teorías abstractas. El código fuente de este software puede ser consultado en el repositorio de github del trabajo: <https://github.com/JAntonioVR/Polinomios-Ortogonales/tree/main/software>.

Palabras clave

Polinomios Ortogonales, Polinomios Ortogonales Clásicos, Cadenas de Markov, Procesos de Nacimiento y Muerte, Fórmula de Representación, Análisis Matemático, Análisis Probabilístico, Procesos Estocásticos, Software Matemático.

Approximation in mathematics has always been one of main alternatives to face the difficulty of obtaining exact solutions to some problems. In this context we can find the orthogonal polynomials, which can be used for, inter alia, approximate functions or calculate integrals and transforms.

On the other hand, the interest and the need to predict future events has caused the origin of many probabilistic models, among which we highlight the Markov chains. In particular, Markov chains and processes are state models, i.e., they try to predict the future situation of a system starting from a very limited information.

It looks like we have been talking about two very different mathematical areas so far, because both disciplines has distinct origin and aims. Nevertheless, there exist such a close relation between orthogonal polynomials and Markov chains. This relation lies in the birth-and-death processes, which are a particular case of Markov chains about which the mathematicians Samuel Karlin (1994–2007) and James McGregor (1926–1991) wrote.

In this final master’s project we will start studying orthogonal polynomials and Markov chains independently in order to finish explaining the relation between them. Thus, this document has been structured in four chapters:

1. **Introduction to Orthogonal Polynomials:** The main concepts and results about orthogonal polynomials theory are explained. The base notation is also established and some clarifying examples are exposed.
2. **Classic Orthogonal Polynomials:** The known as ‘classic’ orthogonal polynomials are presented. They are particular families of polynomials which share some interesting properties that characterize them.
3. **Markov Chains:** Once the orthogonal polynomials theory has been introduced, Markov chains and their behavior are studied in a fully probabilistic way, showing several examples.
4. **Orthogonal Polynomials and Birth-and-Death Processes:** Having the two preliminary theories explained, birth-and-death processes are defined and the relation between these probabilistic models and orthogonal polynomials are deduced.

In addition, throughout this document we show some examples for which mathematical software tools were used to illustrate the theoretical concepts in a practical way. Through the use of **SageMath** and **R**, some symbolic and numeric calculations, graphic renders, temporary measures and comparisons have been made. The source code of this software can be found in the project's github repository: <https://github.com/JAntonioVR/Polinomios-Ortogonales/tree/main/software>.

Keywords

Orthogonal Polynomials, Classic Orthogonal Polynomials, Markov Chains, Birth-and-Death Processes, Representation Formula, Mathematical Analysis, Probabilistic Analysis, Stochastic Processes, Mathematical Software.

Introducción	13
Objetivos	15
1. Introducción a los Polinomios Ortogonales	17
1.1. Introducción	17
1.2. Ortogonalidad y función peso	18
1.3. Funcional de momentos	19
1.4. Estandarizaciones de las SPO	23
1.5. Existencia de SPO	24
1.6. La relación de recurrencia a tres términos	26
1.7. Propiedades de los ceros	30
2. Polinomios Ortogonales Clásicos	35
2.1. La ecuación de Pearson	35
2.2. Deducción de las familias de polinomios ortogonales clásicos	38
2.2.1. Caso I: Polinomios de Hermite	39
2.2.2. Caso II: Polinomios de Laguerre	39
2.2.3. Caso III: Polinomios de Jacobi	40
2.3. Caracterizaciones	40
2.3.1. Ortogonalidad de las derivadas	41
2.3.2. Ecuación diferencial hipergeométrica	42
2.3.3. Fórmula de Rodrigues	44
2.4. Polinomios de Hermite, Laguerre y Jacobi	51
2.4.1. Polinomios de Hermite	51
2.4.2. Polinomios de Laguerre	52
2.4.3. Polinomios de Jacobi	53
3. Cadenas de Markov	55
3.1. Cadenas de Markov discretas. Matriz de transición	55
3.2. Distribuciones transitorias	58

3.3. Comportamiento límite	62
3.3.1. Clasificación de los estados de una cadena de Markov	62
3.3.2. Teoremas límite	64
4. Polinomios Ortogonales y Procesos de Nacimiento y Muerte	69
4.1. Procesos de nacimiento y muerte	69
4.2. La fórmula de representación de Karlin-McGregor	71
4.3. Análogos continuos de la fórmula de representación	78
4.3.1. Procesos de Markov continuos	78
4.3.2. Procesos de nacimiento y muerte continuos	79
4.3.3. La fórmula de representación de Karlin y McGregor continua	80
Conclusiones	83
Bibliografía	83
Apéndices	89
A. Tablas sobre Polinomios Ortogonales clásicos	89
A.1. Relación de recurrencia a tres términos	89
A.2. Ecuación de Pearson	90
A.3. Ecuación diferencial hipergeométrica	90
A.4. Fórmulas de Rodrigues	90
B. El teorema espectral	91

La introducción

OBJETIVOS

Los objetivos

1.1. Introducción

La *Ortogonalidad* es una propiedad que a menudo se ha relacionado con la geometría, siendo común pensar en la analogía con la *perpendicularidad*. Por ejemplo, es claro que en el plano vectorial \mathbb{R}^2 se tiene que los vectores $(1, 0)$ y $(0, 1)$ son ortogonales, y es que estos forman un ángulo recto, y por ello también se dice que son perpendiculares. Sin embargo, este hecho no es más que una consecuencia de la ortogonalidad, y es que en el espacio vectorial \mathbb{R}^d , y en particular en \mathbb{R}^2 , dos vectores se dicen ortogonales si, al dotar a \mathbb{R}^d de un producto escalar $\langle \cdot, \cdot \rangle : \mathbb{R}^d \times \mathbb{R}^d \longrightarrow \mathbb{R}$, el resultado de operar estos dos vectores es 0.

El producto escalar mayormente utilizado en \mathbb{R}^d es el usual, el cual, si $u = (u_1, \dots, u_d)$ y $v = (v_1, \dots, v_d)$ se define como

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathbb{R}^d \times \mathbb{R}^d &\longrightarrow \mathbb{R} \\ (u, v) &\longmapsto \langle u, v \rangle = u \cdot v^t = \sum_{i=1}^d u_i \cdot v_i. \end{aligned}$$

Y dos vectores $u, v \in \mathbb{R}^d$ se dicen ortogonales siempre que $\langle u, v \rangle = 0$.

Sin embargo, en realidad la ortogonalidad va mucho más allá de \mathbb{R}^d y del producto escalar usual. Se puede utilizar cualquier espacio vectorial siempre que esté dotado de un producto escalar. En particular, dado $\Omega \subseteq \mathbb{R}$ pensemos en el espacio de Lebesgue $L^1(\Omega)$, el cual está dotado del producto vectorial

$$\begin{aligned} \langle \cdot, \cdot \rangle : L^1 \times L^1 &\longrightarrow \mathbb{R} \\ (f, g) &\longmapsto \langle f, g \rangle = \int_{\Omega} f(x) \cdot g(x) dx. \end{aligned}$$

Por ejemplo, si tomamos $\Omega = [0, \pi]$, tenemos que las funciones $\cos(n\theta), \cos(m\theta)$ ($n, m \in \mathbb{N}_0$) son ortogonales siempre que $n \neq m$, pues

$$\int_0^\pi \cos(n\theta) \cos(m\theta) d\theta = 0 \quad (n \neq m). \quad (1.1)$$

Si hacemos el cambio de variable $x = \cos(\theta)$, tenemos que $dx = -\sin(\theta)d\theta = \sin(-\theta)d\theta$, por lo que aplicando que $\sin^2(-\theta) + \cos^2(-\theta) = 1$ y que $\cos(-\theta) = \cos(\theta)$, tenemos que la ecuación (1.1) se expresa como

$$\int_{-1}^1 T_n(x)T_m(x)(1-x^2)^{-\frac{1}{2}}dx = 0 \quad (n \neq m). \quad (1.2)$$

donde hemos denotado $T_n(x) = \cos(n\theta) = \cos(n \arccos(\theta))$ con $-1 \leq x \leq 1$.

Y tenemos que $T_0 = 1, T_1 = \cos(\theta) = x$, y aplicando identidades trigonométricas se puede deducir que $T_2 = 2x^2 - 1, T_3 = 4x^3 - 3x$, etc.

Por lo tanto, si definimos en $\mathbb{P}[x]$ el producto escalar

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathbb{P}[x] \times \mathbb{P}[x] &\longrightarrow \mathbb{R} \\ (p, q) &\longmapsto \langle p, q \rangle = \int_{-1}^1 p(x)q(x)\rho(x)dx, \end{aligned}$$

con $\rho(x) = (1-x^2)^{-\frac{1}{2}}$, tenemos que las funciones (polinomios) T_n , $n \in \mathbb{N}_0$ son ortogonales entre sí en el espacio $(\mathbb{P}[x], \langle \cdot, \cdot \rangle)$ siempre que $n \neq m$.

De acuerdo a este ejemplo, decimos que los polinomios T_n son *ortogonales*, o que la sucesión $\{T_n\}$ es una *Sucesión de Polinomios Ortogonales* con respecto a la *función peso* $\rho(x) = (1-x^2)^{-\frac{1}{2}}$ en el intervalo $(-1, 1)$. Los polinomios T_n recién presentados son los *Polinomios de Tchebichef de tipo I*, y conforman nuestra primera familia de polinomios ortogonales conocida.

1.2. Ortogonalidad y función peso

En la sección 1.1 pudimos ampliar el concepto generalmente conocido de ortogonalidad, restringido a espacios como \mathbb{R}^d , a otros tipos de espacios. Además, introdujimos la primera familia de polinomios ortogonales: los polinomios de Tchebichef de tipo I. En esta sección daremos definiciones más formales y genéricas sobre la ortogonalidad, basándonos en los textos de [Álv03, Ch. II], [Chi11, Ch. I] y [Sze75, Ch. II, III].

Sea μ una función no decreciente y no constante definida en un intervalo (a, b) tal que si $a = -\infty$ entonces $\lim_{x \rightarrow -\infty} \mu(x) > -\infty$ y si $b = \infty$ entonces $\lim_{x \rightarrow \infty} \mu(x) < \infty$. Se define el espacio de funciones $L_\mu^p[a, b]$ como el conjunto de funciones tales que

$$\int_a^b |f(x)|^p d\mu(x) < +\infty$$

En $L_\mu^2[a, b]$, se define un producto escalar:

$$\begin{aligned} \langle \cdot, \cdot \rangle : L_\mu^2[a, b] \times L_\mu^2[a, b] &\longrightarrow \mathbb{R} \\ (f, g) &\longmapsto \langle f, g \rangle = \int_a^b f(x)g(x)d\mu(x). \end{aligned} \quad (1.3)$$

A partir del espacio $L_\mu^2[a, b]$ podemos dar una definición de ortogonalidad.

Definición 1.1 (Ortogonalidad). Fijada una función μ no decreciente y no constante definida en un intervalo (a, b) tal que si $a = -\infty$ entonces $\lim_{x \rightarrow -\infty} \mu(x) > -\infty$ y si $b = \infty$ entonces $\lim_{x \rightarrow \infty} \mu(x) < \infty$ y dadas $f, g \in L^2_\mu[a, b]$, se dice que las funciones **f y g son ortogonales** o que **f es ortogonal a g** respecto a la distribución $d\mu$ si

$$\langle f, g \rangle = 0. \quad (1.4)$$

No obstante, en la mayoría de los casos y por simplicidad en lugar de utilizar una función μ y su medida inducida, si μ es absolutamente continua podemos reescribir (1.3) como una integral de Lebesgue

$$\langle f, g \rangle = \int_a^b f(x)g(x)\rho(x)dx, \quad (1.5)$$

siendo ρ una función medible no negativa tal que $0 < \int_a^b \rho(x)dx < \infty$ denominada **función peso**.

Definición 1.2 (Sucesión de Polinomios Ortogonales respecto a una función peso). Dada una función peso ρ , si existe una sucesión de polinomios $\{P_n\}$ con P_n de grado n tal que

$$\langle P_n, P_m \rangle = \int_a^b P_n(x)P_m(x)\rho(x)dx = 0 \quad (n \neq m)$$

entonces decimos que $\{P_n\}$ es una **Sucesión de Polinomios Ortogonales (SPO) respecto a la función peso $\rho(x)$** en el intervalo (a, b) .

1.3. Funcional de momentos

Tenemos ya por tanto una definición rigurosa de lo que es la ortogonalidad de funciones de $L^2_\mu[a, b]$. Podemos reescribir esta propiedad mediante el uso de funcionales lineales aprovechando la linealidad de la integral. Se define el funcional \mathcal{L} como

$$\begin{aligned} \mathcal{L} : L^2_\mu[a, b] &\longrightarrow \mathbb{C} \\ f &\longmapsto \mathcal{L}[f] = \int_a^b f(x)d\mu(x). \end{aligned} \quad (1.6)$$

Por tanto, vemos que la propiedad de ortogonalidad (1.4) es equivalente a $\mathcal{L}[f \cdot g] = 0$.

Observación 1.1. El funcional \mathcal{L} es lineal, pues por la linealidad del operador integral se tiene que

$$\mathcal{L}[af + bg] = a\mathcal{L}[f] + b\mathcal{L}[g], \quad (1.7)$$

para cualesquiera $a, b \in \mathbb{C}$ y $f, g \in L^2_\mu[a, b]$.

Ejemplo 1. $\{T_n\}$, la sucesión de polinomios de Tchebichef de tipo I, es una Sucesión de Polinomios Ortogonales respecto de la función peso $\rho(x) = (1 - x^2)^{-\frac{1}{2}}$ en el intervalo $(-1, 1)$. También podemos establecer la ortogonalidad de los respectivos $\{T_n\}$ no mediante un producto escalar sino a través del funcional

$$\mathcal{L}[f] := \int_{-1}^1 f(x)(1 - x^2)^{-\frac{1}{2}}dx \quad (1.8)$$

Por lo que podemos definir ortogonalidad indistintamente mediante un producto escalar $\langle \cdot, \cdot \rangle$ o mediante un funcional integral \mathcal{L} .

A partir de este momento consideraremos el espacio vectorial de los polinomios \mathbb{P} . Denotaremos como \mathbb{P}_n al subespacio de \mathbb{P} formado por los polinomios de grado menor o igual que n .

Definición 1.3 (Momentos de un funcional). Dado un funcional \mathcal{L} , definimos los **momentos** del funcional, y los denotamos con μ_n , como

$$\mu_n = \mathcal{L}[x^n], \quad n \in \mathbb{N}_0. \quad (1.9)$$

Gracias a los momentos de un funcional y considerando que cualquier polinomio de grado n puede escribirse como $p(x) = \sum_{k=0}^n c_k x^k$ podemos entonces combinar (1.7) y (1.9) para escribir la acción de un funcional \mathcal{L} sobre $p(x)$ como

$$\mathcal{L}[p] = \mathcal{L} \left[\sum_{k=0}^n c_k x^k \right] = \sum_{k=0}^n c_k \mu_k$$

Por lo que, de esta forma, únicamente conociendo la sucesión de momentos $\{\mu_n\}$ podríamos conocer el resultado de aplicar \mathcal{L} a cualquier polinomio. Por tanto, es posible dar una nueva definición de ortogonalidad respecto a funcionales lineales que en este caso estén definidos no mediante una integral como en el caso de (1.6), sino a partir de una sucesión de momentos $\{\mu_n\}$.

Definición 1.4 (Funcional de momentos). Dada una sucesión de números complejos $\{\mu_n\}$, diremos que \mathcal{L} es un **funcional de momentos** determinado por la sucesión $\{\mu_n\}$, donde μ_n se denomina **momento de orden n**, si \mathcal{L} es lineal en \mathbb{P} y $\mathcal{L}[x^n] = \mu_n$, $n \in \mathbb{N}_0$

Daremos ahora una nueva definición de SPO, en este caso con respecto a un funcional lineal y no respecto a una función peso como hicimos en la definición 1.2.

Definición 1.5 (Sucesión de Polinomios Ortogonales respecto a un funcional lineal). Dado un funcional lineal \mathcal{L} definido como en (1.6), una sucesión de polinomios $\{P_n\}$ es una **Sucesión de Polinomios Ortogonales (SPO) respecto al funcional \mathcal{L}** si:

1. P_n es de grado n .
2. $\mathcal{L}[P_n P_m] = 0$ si $n \neq m$, $n, m \in \mathbb{N}_0$.
3. $\mathcal{L}[P_n^2] \neq 0 \quad \forall n \in \mathbb{N}_0$.

La sucesión $\{P_n\}$ se llamará **ortonormal** si $\mathcal{L}[P_n^2] = 1 \quad \forall n \in \mathbb{N}_0$.

La sucesión $\{P_n\}$ se llamará **Sucesión de Polinomios Ortogonales Mónicos (SPOM)** si P_n es mónico para cada n , es decir, si $P_n(x) = x^n + a_{n-1}x^{n-1} + \dots + a_0$.

En general, las condiciones (2) y (3) se suelen abreviar como

$$\mathcal{L}[P_n P_m] = K_n \delta_{mn}, \quad K_n \neq 0,$$

donde δ_{mn} denota la función delta de Kronecker.

En el caso de que $d\mu(x) = \rho(x)dx$, es decir, en el caso de trabajar con funciones peso, la condición (3) se satisface de manera automática.

Ejemplo 2. Extraído de [Chi11, Capítulo 1, sección 1] Consideramos

$$P_n(x) = \sum_{k=0}^n \binom{x}{k} \frac{(-a)^{n-k}}{(n-k)!}.$$

$\{P_n\}$ son los llamados *Polinomios de Charlier*. Como

$$\binom{x}{n} = \frac{1}{n!} x(x-1) \cdots (x-n+1)$$

tenemos que $P_n(x)$ es un polinomio de grado n . En [Chi11] pueden encontrarse los cálculos rigurosos mediante los cuales, si definimos

$$\mathcal{L}[x^n] = \sum_{k=0}^{\infty} k^n \frac{a^k}{k!}, \quad n \in \mathbb{N}_0.$$

y extendemos \mathcal{L} a \mathbb{P} por linealidad, tenemos que

$$\mathcal{L}[P_m(x)P_n(x)] = \frac{e^a a^n}{n!} \delta_{mn}, \quad m, n \in \mathbb{N}_0.$$

Al estar definido el funcional mediante sumatorias y no mediante integrales podría parecer que, aunque $\{P_n\}$ satisfaga la definición 1.5 mediante un funcional de momentos, no satisface la ortogonalidad respecto a (1.6). Sin embargo, si denotamos como ϕ a una función escalonada que es constante en cada intervalo $(-\infty, 0)$ y $(k, k+1)$ con $k \in \mathbb{N}_0$ y tiene saltos de magnitud $\frac{a^k}{k!}$ en cada k , entonces utilizando la integral de Riemann-Stieltjes los polinomios $\{P_n\}$ verifican

$$\int_{-\infty}^{\infty} P_m(x)P_n(x)d\phi(x) = \frac{e^a a^n}{n!} \delta_{mn}, \quad m, n \in \mathbb{N}_0. \quad (1.10)$$

Evidentemente ϕ no es una función absolutamente continua, por lo que no se puede escribir (1.10) en términos de (1.5).

Observación 1.2. No cualquier sucesión $\{\mu_n\}$ da lugar a una SPO. Por ejemplo, si $\mu_0 = \mu_1 = \mu_2 = 1$, tendríamos que

$$P_0(x) = a \neq 0, \quad P_1(x) = bx + c, \quad b \neq 0.$$

Por la propiedad (2) de la definición 1.5

$$\mathcal{L}[P_0(x)P_1(x)] = a(b\mu_1 + c\mu_0) = a(b+c),$$

luego $b = -c$, y en este caso

$$\mathcal{L}[P_1(x)^2] = b^2(\mu_2 - 2\mu_1 + \mu_0) = 0,$$

lo cual contradice la propiedad (3).

Gracias a la observación 1.2 sabemos que no es válida cualquier sucesión de momentos para encontrar una SPO, por lo que necesitamos además probar y dar condiciones para la existencia de la susodicha SPO.

Previamente, introduciremos un resultado para la caracterización de SPO respecto a un funcional \mathcal{L} .

Teorema 1.1. Sea \mathcal{L} un funcional de momentos y $\{P_n\}$ una sucesión de polinomios. Las siguientes afirmaciones son equivalentes:

1. $\{P_n\}$ es una SPO respecto al funcional \mathcal{L} .
2. $\mathcal{L}[\pi(x)P_n(x)] = 0$ para cualquier polinomio $\pi(x)$ de grado $m < n$ y $\mathcal{L}[\pi(x)P_n(x)] \neq 0$ si $\pi(x)$ tiene grado n .
3. $\mathcal{L}[x^m P_n(x)] = K_n \delta_{nm}$, con $K_n \neq 0$, $m = 0, 1, \dots, n$

Demostración. Probaremos $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (1)$.

■ $(1) \Rightarrow (2)$

Supongamos que $\{P_n\}$ es una SPO respecto al funcional \mathcal{L} . Como P_k tiene grado k , entonces $\{P_0, \dots, P_m\}$ forma una base de \mathbb{P}_m . Por lo que si $\pi(x)$ es un polinomio de grado m , existen constantes c_1, \dots, c_m , con $c_m \neq 0$ tales que $\pi(x) = \sum_{i=0}^m c_i P_i(x)$. Como \mathcal{L} es lineal,

$$\mathcal{L}[\pi(x)P_n(x)] = \sum_{k=0}^m c_k \mathcal{L}[P_k(x)P_n(x)] = \begin{cases} 0 & \text{si } m < n \\ c_n \mathcal{L}[P_n^2(x)] & \text{si } m = n \end{cases}$$

■ $(2) \Rightarrow (3)$ es trivial, basta con utilizar $\pi(x) = x^m$ y aplicar (2).

■ $(3) \Rightarrow (1)$ también es trivial, pudiendo reconstruir $P_m(x)$ por linealidad.

□

En la prueba de este último teorema, hemos hecho uso de que mediante una SPO podemos obtener bases de \mathbb{P}_n . Sacaremos provecho de este hecho en el siguiente resultado.

Teorema 1.2. Sea $\{P_n\}$ una SPO respecto a un funcional lineal \mathcal{L} . Entonces, cualquier polinomio $\pi(x)$ de grado n se puede expresar en la base $\{P_0, \dots, P_n\}$ de la forma

$$\pi(x) = \sum_{k=0}^n c_k P_k(x), \quad \text{con } c_k = \frac{\mathcal{L}[\pi(x)P_k(x)]}{\mathcal{L}[P_k^2(x)]}.$$

Demostración. Ya es claro que si $\pi(x)$ es un polinomio de grado n , entonces $\pi(x) = \sum_{k=0}^n c_k P_k(x)$. Por tanto, multiplicando ambos miembros de la igualdad por $P_m(x)$ y aplicando el funcional \mathcal{L} se obtiene

$$\mathcal{L}[\pi(x)P_m(x)] = \sum_{k=0}^n c_k \mathcal{L}[P_k(x)P_m(x)] = c_m \mathcal{L}[P_m^2(x)],$$

donde hemos aplicado la ortogonalidad. De esta igualdad se deduce claramente la expresión de c_k .

□

Este resultado nos proporciona una consecuencia muy importante: la unicidad salvo constante multiplicativa de las SPO respecto a un funcional.

Corolario 1.3. Sea $\{P_n\}$ una SPO respecto a un funcional lineal \mathcal{L} . Entonces cada P_n está unívocamente determinado salvo constante multiplicativa no nula. Esto es, si $\{Q_n\}$ es otra SPO respecto a \mathcal{L} , entonces existen constantes $c_n \neq 0$ tales que

$$Q_n(x) = c_n P_n(x), \quad n \in \mathbb{N}_0.$$

Demostración. Utilizaremos el teorema 1.2 con $\pi(x) = Q_n(x)$, de modo que

$$Q_n(x) = \sum_{k=0}^n \frac{\mathcal{L}[Q_n(x)P_k(x)]}{\mathcal{L}[P_k^2(x)]} P_k(x)$$

Por el teorema 1.1, $\mathcal{L}[Q_n(x)P_k(x)] = 0$ siempre que $k < n$ y $\mathcal{L}[Q_n(x)P_k(x)] = r_n \neq 0$ si $k = n$, por lo que, tomando $c_n = \frac{r_n}{\mathcal{L}[P_n^2(x)]}$, se tiene $Q_n(x) = c_n P_n(x)$, como queríamos probar. \square

1.4. Estandarizaciones de las SPO

El corolario 1.3 nos afirma que si tenemos una SPO $\{P_n\}$ respecto a un funcional \mathcal{L} , entonces $\{c_n P_n\}$ también es una SPO para cualquier sucesión $\{c_n\}$ de constantes no nulas. Existen varias formas de estandarizar una SPO de forma que esta esté unívocamente determinada a partir de un funcional. En esta sección presentaremos algunas de las más comunes.

Una primera posibilidad es imponer el valor del coeficiente líder de cada polinomio de la sucesión. Por ejemplo, la estandarización más común de los polinomios de Laguerre $L_n^{(\alpha)}$ consiste en obligar a que el término líder a_n verifique $a_n = \frac{(-1)^n}{n!}$.

En particular, podemos exigir que todos los polinomios que conforman la SPO sean mónicos. De esta forma, podemos afirmar que si tenemos una SPO $\{P_n\}$ para un funcional \mathcal{L} , entonces existe una única sucesión de polinomios ortogonales mónicos (SPOM) respecto a \mathcal{L} . De hecho, si denotamos como a_n al coeficiente líder de P_n , entonces los polinomios

$$\hat{P}_n = a_n^{-1} P_n$$

forman una SPOM respecto a \mathcal{L} .

Otra forma de estandarizar una SPO $\{P_n\}$ respecto a un funcional \mathcal{L} es tomando los polinomios

$$p_n = (-1)^{\text{sgn}(a_n)} (\mathcal{L}[P_n^2])^{-1/2} P_n,$$

siempre que $\mathcal{L}[P_n^2] > 0$ y donde $\text{sgn}(a_n)$ denota el signo de a_n . Esta normalización no siempre es posible de aplicar, ya que no todos los funcionales lineales verifican $\mathcal{L}[P_n^2] > 0$, aunque existe una clase de funcionales, los definidos positivos (ver definición 1.8), que aseguran esta condición.

Con esta operación conseguimos que $\mathcal{L}[p_n^2] = 1$, $\forall n \in \mathbb{N}_0$, es decir, la sucesión $\{p_n\}$ es una SPO ortonormal. Por tanto, dada una SPO $\{P_n\}$ para un funcional \mathcal{L} tal que $\mathcal{L}[P_n^2] > 0 \forall n \in \mathbb{N}_0$, existe una única SPO ortonormal $\{p_n\}$ respecto a \mathcal{L} .

Por último, es común exigirle a cada polinomio P_n de la sucesión que verifique $P_n(x_0) = K_n$ para cierto $x_0 \in [a, b]$. Es decir, se exige un valor concreto al evaluar

P_n en un punto del intervalo $[a, b]$, que en ocasiones coincide con alguno de los extremos a, b cuando estos son finitos.

Por ejemplo, los polinomios de Tchebichef de tipo I, ortogonales en $(-1, 1)$ suelen ser presentados verificando $T_n(1) = 1$. Por su parte, los polinomios de Jacobi $P_n^{(\alpha, \beta)}$, que también son ortogonales en $(-1, 1)$, se busca que verifiquen $P_n^{(\alpha, \beta)} = \binom{n+\alpha}{n}$.

En conclusión, no existe una única forma de estandarizar una SPO, sino que en realidad cada familia de polinomios ortogonales acepta una, o en ocasiones varias, estandarizaciones concretas dependiendo de la aplicación que se le dé y del contexto del problema.

1.5. Existencia de SPO

Recordemos que con anterioridad hemos comentado en la observación 1.2 que no cualquier sucesión de números complejos $\{\mu_n\}$ da lugar a una SPO. Es necesario por tanto exigir alguna condición sobre dicha sucesión si realmente queremos asegurar la existencia de la correspondiente SPO.

Para ello, introducimos el siguiente concepto.

Definición 1.6 (Matriz de momentos). Llamaremos **matriz de momentos** a la matriz $(\mu_{i+j})_{i,j=0}^n$, y **determinante de momentos** a su determinante

$$\Delta_n = \det(\mu_{i+j})_{i,j=0}^n = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{vmatrix}. \quad (1.11)$$

Este determinante está directamente relacionado con la existencia de una SPO.

Teorema 1.4. Sea $\{\mu_n\}$ una sucesión de números complejos y sea \mathcal{L} un funcional de momentos cuya sucesión de momentos es $\{\mu_n\}$. \mathcal{L} admite una sucesión de polinomios ortogonales si, y solo si el determinante de momentos $\Delta_n \neq 0 \quad \forall n \in \mathbb{N}_0$.

Demostración. Comprobamos la doble implicación.

\Rightarrow Supongamos que los polinomios $P_n(x) = \sum_{k=0}^n c_{nk}x^k$ conforman una SPO para \mathcal{L} . Fijemos ahora $n \in \mathbb{N}_0$ arbitrario. Por el teorema 1.1 (caracterización de una SPO), obsérvese que la condición de ortogonalidad

$$\mathcal{L}[x^m P_n(x)] = \sum_{k=0}^n c_{nk} \mu_{k+m} = K_n \delta_{nm}, \quad K_n \neq 0, m \leq n \quad (1.12)$$

equivale al sistema

$$\begin{pmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{pmatrix} \begin{pmatrix} c_{n0} \\ c_{n1} \\ \vdots \\ c_{nn} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ K_n \end{pmatrix}. \quad (1.13)$$

Por tanto, si esta sucesión existe, necesariamente este sistema debe tener solución única determinada por K_n . Para ello ha de verificarse que $\Delta_n \neq 0, \quad \forall n \in \mathbb{N}_0$.

⇐ Recíprocamente, supongamos ahora que $\Delta_n \neq 0, \forall n \in \mathbb{N}_0$. En ese caso el sistema (1.13) tiene solución única para cualquier $K_n \neq 0$, por lo que podemos crear polinomios $P_n(x)$ que cumplan (1.12) para cualquier $n \in \mathbb{N}_0$. Por otro lado, aplicando la regla de Cramer, tenemos que

$$c_{nn} = \frac{K_n \Delta_{n-1}}{\Delta_n} \neq 0, \quad n \geq 1, \quad (1.14)$$

expresión que también es válida para $n = 0$ sin más que tomar $\Delta_{-1} = 1$. Por tanto, el polinomio $P_n(x)$ tiene coeficiente líder no nulo, es decir, es de grado n , por lo que $\{P_n(x)\}$ es una SPO para \mathcal{L} .

□

Con esta notación, podemos ponerle nombre a aquellos funcionales para los cuales existe una SPO.

Definición 1.7 (Funcional cuasi-definido). Un funcional \mathcal{L} es **cuasi-definido** si, y solo si admite una sucesión de polinomios ortogonales. Equivalentemente, si $\Delta_n \neq 0, \forall n \in \mathbb{N}_0$.

Un concepto muy relacionado con el de funcional cuasi-definido es el de definido positivo, que introducimos a continuación.

Definición 1.8 (Funcional definido positivo). Un funcional de momentos \mathcal{L} se dice que es **definido positivo** si $\mathcal{L}[\pi] > 0$ para cualquier polinomio π no nulo y no negativo en todo el eje real.

Presentamos ahora un sencillo lema que nos ayudará a encontrar una definición alternativa utilizada con bastante frecuencia.

Lema 1.1. Sea $\pi(x)$ un polinomio no negativo en todo el eje real. Entonces existen polinomios $p(x)$ y $q(x)$ tales que

$$\pi(x) = p^2(x) + q^2(x).$$

Demostración. Si $\pi(x) \geq 0 \forall x \in \mathbb{R}$, entonces $\pi(x)$ es un polinomio cuyos ceros reales tienen multiplicidad par y cuyos ceros complejos son pares conjugados. Por lo que podemos escribir

$$\pi(x) = r^2(x) \prod_{k=1}^m (x - a_k - b_k i)(x - a_k + b_k i),$$

donde $r(x)$ es un polinomio real y $a_k, b_k \in \mathbb{R}$. Si ahora denotamos

$$\prod_{k=1}^m (x - a_k - b_k i) = A(x) + iB(x),$$

siendo $A(x)$ y $B(x)$ polinomios reales, podemos expresar $\pi(x)$ como

$$\pi(x) = r^2(x)(A^2(x) + B^2(x)) = \underbrace{(r(x)A(x))^2}_{p^2(x)} + \underbrace{(r(x)B(x))^2}_{q^2(x)}.$$

□

Una consecuencia directa de este lema es una definición alternativa y equivalente de funcional definido positivo.

Corolario 1.5. *Un funcional de momentos \mathcal{L} es definido positivo si, y solo si $\mathcal{L}[\pi^2(x)] > 0$ para cualquier polinomio $\pi(x)$.*

Finalmente, introducimos la relación entre el concepto de funcional definido positivo y los determinantes (1.11).

Teorema 1.6. *Un funcional \mathcal{L} es definido positivo si, y solo si sus momentos $\{\mu_n\}$ son todos reales y $\Delta_n > 0 \quad \forall n \in \mathbb{N}_0$*

Demostración. La prueba de este resultado puede encontrarse en [Chi11, Teorema 3.4]. \square

Obsérvese que gracias a este teorema podemos entonces afirmar que los funcionales definidos positivos son, en particular, funcionales cuasi-definidos. Esto es, admiten una SPO.

1.6. La relación de recurrencia a tres términos

Las SPO respecto a un funcional siempre cumplen una ecuación que es la denominada ‘relación de recurrencia a tres términos’, en adelante también referida como ‘RRTT’.

Teorema 1.7 (Relación de Recurrencia a Tres Términos). *Sea $\{P_n\}$ una SPO respecto a un funcional lineal \mathcal{L} . Entonces, $\{P_n\}$ satisface la relación de recurrencia*

$$xP_n(x) = \alpha_n P_{n+1}(x) + \beta_n P_n(x) + \gamma_n P_{n-1}(x) \quad \forall n \in \mathbb{N}_0. \quad (1.15)$$

Normalmente se impone que $P_{-1} = 0, P_0 = 1$, por lo que una SPO queda determinada unívocamente a partir de las sucesiones $\{\alpha_n\}, \{\beta_n\}, \{\gamma_n\}$. De hecho, se cumple que $\alpha_n, \gamma_n \neq 0 \quad \forall n \in \mathbb{N}_0$.

Además, si la SPO es mónica, los polinomios $\{P_n\}$ verifican la ecuación equivalente

$$P_{n+1}(x) = (x - \beta_n)P_n - \gamma_n P_{n-1} \quad \forall n \in \mathbb{N}_0. \quad (1.16)$$

Demostración. Como $xP_n(x)$ es un polinomio de grado $n+1$, este puede ser expresado como

$$xP_n(x) = \sum_{k=0}^{n+1} a_{k,n} P_k(x), \quad \text{con } a_{k,n} = \frac{\mathcal{L}[xP_n(x)P_k(x)]}{\mathcal{L}[P_k^2(x)]}.$$

Pero $\mathcal{L}[xP_n(x)P_k(x)] = 0$ para $k = 0, \dots, n-2$, por lo que necesariamente debe cumplirse

$$xP_n(x) = \underbrace{a_{n-1,n}}_{\gamma_n} P_{n-1}(x) + \underbrace{a_{n,n}}_{\beta_n} P_n(x) + \underbrace{a_{n+1,n}}_{\alpha_n} P_{n+1}(x),$$

de donde se deduce (1.15). Si además $P_n(x)$ es mónico, $xP_n(x)$ también lo es, por lo que $a_{n+1,n} = 1$, luego se tiene que

$$xP_n(x) = P_{n+1}(x) + \beta_n P_n(x) + \gamma_n P_{n-1},$$

que es equivalente a (1.16).

Además, $\alpha_n = a_{n+1,n} = \frac{\mathcal{L}[xP_n(x)P_{n+1}(x)]}{\mathcal{L}[P_{n+1}^2(x)]} \neq 0$ porque $\mathcal{L}[xP_n(x)P_{n+1}(x)] \neq 0$ al ser $xP_n(x)$ un polinomio de grado $n+1$ al igual que $P_{n+1}(x)$. Por su parte, $\gamma_n = a_{n-1,n} = \frac{\mathcal{L}[xP_n(x)P_{n-1}(x)]}{\mathcal{L}[P_{n-1}^2(x)]} \neq 0$ porque $\mathcal{L}[xP_n(x)P_{n-1}(x)] \neq 0$ al ser $xP_{n-1}(x)$ un polinomio de grado n al igual que $P_n(x)$. \square

De esta demostración podemos deducir además la expresión de las constantes α_n, β_n y γ_n :

$$\alpha_n = \frac{\mathcal{L}[xP_nP_{n+1}]}{\mathcal{L}[P_{n+1}^2]}, \quad \beta_n = \frac{\mathcal{L}[xP_n^2]}{\mathcal{L}[P_n^2]}, \quad \gamma_n = \frac{\mathcal{L}[xP_nP_{n-1}]}{\mathcal{L}[P_{n-1}^2]}, \quad (1.17)$$

donde la expresión de α_n y la de β_n son válidas para $n \geq 0$ y la de γ_n para $n \geq 1$.

Sin embargo, este cálculo puede complicarse bastante, por lo que presentaremos un algoritmo alternativo. Para ello, desarrollaremos los polinomios como

$$P_n(x) = a_n x^n + b_n x^{n-1} + c_n x^{n-2} + \dots$$

Si sustituimos esta expresión en la RRTT (1.15) e igualamos los coeficientes de los monomios x^{n+1}, x^n y x^{n-1} llegamos a las expresiones

$$\alpha_n = \frac{a_n}{a_{n+1}}, \quad \beta_n = \frac{b_n}{a_n} - \frac{b_{n+1}}{a_{n+1}}, \quad \gamma_n = \frac{1}{a_{n-1}} (c_n - \alpha_n c_{n+1} - \beta_n b_n). \quad (1.18)$$

donde de nuevo la expresión de α_n y la de β_n son válidas para $n \geq 0$ y la de γ_n para $n \geq 1$.

Con las ecuaciones (1.17) podemos construir la SPO utilizando la RR3T si conocemos el funcional, deduciendo simultáneamente los coeficientes de la RR3T y los polinomios que conforman la sucesión. Por su parte, utilizando (1.18) podemos calcular de una manera sencilla los coeficientes de la RR3T sin utilizar el funcional ni integrales, en ocasiones complejas. Sin embargo, en este último caso necesitamos conocer los polinomios P_{n-1}, P_n y P_{n+1} para obtener dichos coeficientes, no pudiendo construir la SPO como sí podíamos hacer si aplicábamos (1.17). Comprobemos este hecho con el siguiente ejemplo.

Ejemplo 3. Vamos a clarificar estas ideas con un ejemplo práctico. En el archivo `software/Hermite.ipynb`¹ se ha implementado, utilizando el software matemático ‘SageMath’, el cálculo de una SPO mediante la RR3T. También se ha realizado una estimación de los tiempos de ejecución para distintos valores de n . En este ejemplo únicamente se recogen los resultados arrojados por el software, pero en dicho archivo se puede consultar e incluso modificar el código fuente.

En particular, se ha trabajado con los polinomios de Hermite, (véase sección 2.2.1) que son aquellos que conforman la SPO con respecto a la función peso $\rho(x) = e^{-x^2}$ en todo \mathbb{R} . Por tanto, tenemos el funcional

$$\mathcal{L}[f] = \int_{-\infty}^{+\infty} f(x) e^{-x^2} dx.$$

¹Puede ser consultado en la URL <https://github.com/JAntonioVR/Polinomios-Ortogonales/blob/main/software/Hermite.ipynb>

Calculemos la SPOM correspondiente al funcional \mathcal{L} . Si tomamos $P_{-1} = 0, P_0 = 1$, podemos calcular a partir de (1.17) las constantes β_n, γ_n para $n \geq 1$ y mediante (1.16) obtenemos los polinomios P_n para $n \geq 1$, véase la tabla 1.1.

n	$P_n(x)$	Tiempo de cálculo (s)
0	1	1.1897×10^{-5}
1	x	0.0133
2	$x^2 - \frac{1}{2}$	0.0449
3	$x^3 - \frac{3}{2}x$	0.1073
4	$x^4 - 3x^2 + \frac{3}{4}$	0.2161
5	$x^5 - 5x^3 + \frac{15}{4}x$	0.4162
6	$x^6 - \frac{15}{2}x^4 + \frac{45}{4}x^2 - \frac{15}{8}$	0.6752
7	$x^7 - \frac{21}{2}x^5 + \frac{105}{4}x^3 - \frac{105}{8}x$	1.2009
8	$x^8 - 14x^6 + \frac{105}{2}x^4 - \frac{105}{2}x^2 + \frac{105}{16}$	2.0665
9	$x^9 - 18x^7 + \frac{189}{2}x^5 - \frac{315}{2}x^3 + \frac{945}{16}x$	3.5725
10	$x^{10} - \frac{45}{2}x^8 + \frac{315}{2}x^6 - \frac{1575}{4}x^4 + \frac{4725}{16}x^2 - \frac{945}{32}$	5.5059

Tabla 1.1: 10 primeros polinomios de Hermite mónicos

Sin embargo, comprobamos que este método es lento, pues los tiempos de cálculo² se incrementan considerablemente con n . De hecho, para $n = 15$, que es un valor aún bajo y cercano a 10, el tiempo de ejecución ya es de 58.9024 segundos, casi un minuto.

Por otro lado, existe un resultado recíproco al teorema 1.7. Este resultado es el conocido como teorema de Favard.

Teorema 1.8 (Favard). *Sean $\{\beta_n\}$ y $\{\gamma_n\}$ dos sucesiones de números reales y sea $\{P_n\}$ una sucesión de polinomios mónicos definida mediante la relación (1.16):*

$$P_{n+1}(x) = (x - \beta_n)P_n - \gamma_n P_{n-1} \quad \forall n \in \mathbb{N}_0.$$

con $P_{-1} = 0$ y $P_0 = 1$. Entonces existe un único funcional de momentos \mathcal{L} tal que

$$\mathcal{L}[1] = \gamma_0, \quad \mathcal{L}[P_n P_m] = \delta_{n,m} K_n, \quad K_n \neq 0.$$

Este funcional es cuasi-definido y $\{P_n\}$ es su correspondiente SPOM si, y solo si $\gamma_n \neq 0 \forall n \in \mathbb{N}_0$. Además, \mathcal{L} es definido positivo si y solo si $\gamma_n > 0 \forall n \in \mathbb{N}_0$.

Demostración. Definamos el funcional \mathcal{L} por inducción en \mathbb{P} . Sea

$$\mathcal{L}[1] = \mathcal{L}[P_0] = \mu_0 := \gamma_0,$$

$$\mathcal{L}[P_n] = \mathcal{L}[1 \cdot P_n] = \mathcal{L}[P_0 P_n] := 0 \quad \forall n = 1, 2, \dots$$

Así, aplicando (1.16) podemos calcular los momentos del funcional. Si tomamos $n = 0$ en (1.16) obtenemos

$$P_1(x) = x - \beta_0,$$

²Estos tiempos han sido obtenidos mediante la media de los tiempos de ejecución de 10 ejecuciones independientes.

luego, como $\mathcal{L}[P_1] = 0$:

$$\begin{aligned} 0 = \mathcal{L}[P_1] &= \mathcal{L}[x - \beta_0] \\ &= \mu_1 - \beta_0 \mathcal{L}[1] \\ &= \mu_1 - \beta_0 \gamma_0, \end{aligned}$$

de donde $\mu_1 = \beta_0 \gamma_0$. Si $n = 1$ en (1.16) obtenemos

$$P_2(x) = (x - \beta_1)P_1 - \gamma_1,$$

por tanto,

$$\begin{aligned} 0 = \mathcal{L}[P_2] &= \mathcal{L}[(x - \beta_1)P_1 - \gamma_1] \\ &= \mathcal{L}[(x - \beta_1)(x - \beta_0) - \gamma_1] \\ &= \mu_2 - (\beta_0 + \beta_1)\mu_1 + (\beta_1\beta_0 - \gamma_1)\gamma_0, \end{aligned}$$

de donde $\mu_2 = (\beta_0 + \beta_1)\mu_1 - (\beta_1\beta_0 - \gamma_1)\gamma_0$. Continuando este proceso se pueden obtener todos los momentos μ_n , y una vez se tienen todos los momentos podemos extender el funcional por linealidad.

Utilizando (1.15) sucesivamente obtenemos que

$$x^k P_n = \sum_{i=n-k}^{n+k} d_{n,i} P_i(x),$$

y aplicando que $\mathcal{L}[P_n] = 0$ se deduce que $\mathcal{L}[x^k P_n] = 0$ para todo $k = 1, \dots, n-1$. Comprobemos qué ocurre cuando $k = n$. De nuevo, por (1.15):

$$\mathcal{L}[x^n P_n] = \mathcal{L}[x^{n-1}(P_{n+1} + \beta_n P_n + \gamma_n P_{n-1})] = \mathcal{L}[\gamma_n x^{n-1} P_{n-1}] = \gamma_n \mathcal{L}[x^{n-1} P_{n-1}].$$

Aplicando este razonamiento sucesivamente llegamos a que

$$\mathcal{L}[x^n P_n] = \gamma_n \gamma_{n-1} \cdots \gamma_0 \quad \forall n \in \mathbb{N}_0. \quad (1.19)$$

Por tanto, el funcional es cuasi-definido y $\{P_n\}$ es su SPOM si, y solo si $\gamma_n \neq 0 \quad \forall n \in \mathbb{N}_0$.

Por último, véamos que \mathcal{L} es definido positivo si y solo si $\gamma_n > 0 \quad \forall n \in \mathbb{N}_0$. Sea $\pi(x)$ un polinomio y m su grado. Podemos entonces escribir $\pi(x) = \sum_{i=1}^m c_i P_i(x)$. De esta forma

$$\mathcal{L}[\pi^2(x)] = \mathcal{L}\left[\sum_{i=1}^m c_i^2 P_i^2(x)\right] = \sum_{i=1}^m c_i^2 \mathcal{L}[P_i^2(x)],$$

donde hemos utilizado la ortogonalidad de $\{P_n\}$. Por tanto, la condición $\mathcal{L}[\pi^2(x)] > 0$ es equivalente a $\mathcal{L}[P_n^2(x)] > 0 \quad \forall n \in \mathbb{N}_0$. Aplicando de nuevo la ortogonalidad de la sucesión, tenemos que

$$\mathcal{L}[P_n^2(x)] = \mathcal{L}[P_n \cdot P_n] = \mathcal{L}[x^n P_n],$$

que por (1.19) es positivo si y solo si $\gamma_n \gamma_{n-1} \cdots \gamma_0 > 0$, que es equivalente a que $\gamma_n > 0 \quad \forall n \in \mathbb{N}_0$. \square

Observación 1.3. Nótese que el teorema de Favard está descrito partiendo de una sucesión de polinomios ortogonales *mónicos*. Sin embargo, una vez es calculado el funcional, se puede utilizar cualquier otra estandarización sin cambiar el funcional.

Si partimos de una sucesión de polinomios mónicos $\{P_n\}$ y dos sucesiones de números reales $\{\beta_n\}$ y $\{\gamma_n\}$ verificando las condiciones del teorema de Favard, entonces existe un funcional \mathcal{L} para el cual $\{P_n\}$ es una SPOM. Si $\{Q_n\} = \{c_n P_n\}$ es otra SPO para \mathcal{L} , podemos sustituir $P_n = \frac{1}{c_n} Q_n$ en (1.17) para obtener la relación

$$xQ_n(x) = \frac{c_n}{c_{n+1}}Q_{n+1} + \beta_n Q_n + \frac{\gamma_n}{c_{n-1}}Q_{n-1}.$$

1.7. Propiedades de los ceros

Los ceros o raíces de los polinomios ortogonales tienen propiedades y comportamientos que guardan cierta regularidad cuando el funcional que los define es definido positivo. En esta sección conoceremos estas peculiaridades de los ceros. Comencemos con una definición que refina en cierto modo la definición de funcional definido positivo.

Definición 1.9 (Funcional definido positivo en un conjunto). Sea $E \subseteq \mathbb{R}$. Decimos que un funcional \mathcal{L} es **definido positivo en E** si $\mathcal{L}[\pi] > 0$ para cualquier polinomio $\pi(x)$ real, no negativo y no idénticamente nulo en E . En este contexto, se dice que E es un **soporte** de \mathcal{L} .

Por ejemplo, el funcional de momentos (1.8) de los polinomios de Tchebichef de tipo I es definido positivo en $(-1, 1)$, el de los polinomios de Charlier definido en el ejemplo 2 lo es en \mathbb{N} y el de los polinomios de Hermite en \mathbb{R} . Respecto a este último caso, es claro que ser ‘definido positivo en \mathbb{R} ’ equivale a ser ‘definido positivo’ según la definición 1.8.

A partir de este concepto y los siguientes teoremas podremos conocer la dinámica que siguen los ceros de los polinomios ortogonales. En adelante, \mathcal{L} es un funcional definido positivo y $\{P_n(x)\}$ su correspondiente SPO.

Teorema 1.9. Sea I un intervalo que es un soporte de \mathcal{L} . Entonces los ceros de P_n ($n \geq 1$) son todos simples, reales y se encuentran en el interior de I (denotado en adelante como $\overset{\circ}{I}$).

Demostración. Supongamos que $P_0 = 1$, entonces $\mathcal{L}[P_n] = \mathcal{L}[P_0 P_n] = 0$, $n \geq 1$. Por tanto, P_n tiene que cambiar de signo al menos una vez en $\overset{\circ}{I}$. Esto es, P_n tiene al menos un cero con multiplicidad impar en $\overset{\circ}{I}$. Sean $x_1, \dots, x_k \in \overset{\circ}{I}$ los ceros de multiplicidad impar de P_n que están en $\overset{\circ}{I}$. Consideramos entonces el polinomio

$$q(x) = (x - x_1) \cdots (x - x_k).$$

Entonces $q(x)P_n(x)$ es un polinomio que no tiene ceros de multiplicidad impar en $\overset{\circ}{I}$, por lo que $q(x)P_n(x) \geq 0 \forall x \in I$. Por tanto, $\mathcal{L}[q(x)P_n(x)] > 0$, pero esto contradice el teorema 1.1 salvo que $k = n$, luego necesariamente $P_n(x)$ tiene n ceros simples en $\overset{\circ}{I}$. \square

A partir de este teorema, podemos denotar los ceros del polinomio $P_n(x)$ como

$$x_{n1} < x_{n2} < \cdots < x_{nn}, \quad n \geq 1.$$

Observación 1.4. Si tomamos P_n con coeficiente líder positivo, entonces tenemos que

1. $P_n(x) > 0 \quad \forall x > x_{nn}$ y
2. $\text{sgn} P_n(x) = (-1)^n \quad \forall x < x_{n1}$.

Observación 1.5. P'_n tiene exactamente un cero en cada intervalo $(x_{n,k}, x_{n,k+1})$, por lo que $P'_n(x_{nk})$ alterna su signo con $k = 1, \dots, n$. Como el signo del coeficiente líder de $P'_n(x)$ también es positivo, concluimos que

$$\text{sgn} P'_n(x_{nk}) = (-1)^{n-k}, \quad k = 1, \dots, n. \quad (1.20)$$

Estas dos observaciones relacionan en cierto modo los ceros de P_n con el comportamiento de P'_n . El siguiente lema relaciona directamente la SPO $\{P_n\}$ con la sucesión de las derivadas de los polinomios ortogonales $\{P'_n\}$.

Lema 1.2. Si \mathcal{L} es un funcional definido positivo y $\{P_n\}$ es su SPO, entonces se verifica la desigualdad

$$P'_{n+1}(x)P_n(x) - P'_n(x)P_{n+1}(x) > 0, \quad \forall x \in \mathbb{R}. \quad (1.21)$$

Esta desigualdad es una consecuencia de la conocida *Identidad de Christoffel-Darboux* (véase [Chi11, Theorem 4.5]), y nos será de gran utilidad para probar el siguiente teorema.

Teorema 1.10 (Serparación de los ceros). *Los ceros de P_n y de P_{n+1} están mutuamente separados. Es decir,*

$$x_{n+1,i} < x_{n,i} < x_{n+1,i+1}, \quad i = 1, \dots, n.$$

Demostración. A partir de la desigualdad (1.21), si evaluamos en $x = x_{n+1,k}$ se tiene que

$$P'_{n+1}(x_{n+1,k})P_n(x_{n+1,k}) > 0, \quad k = 1, \dots, n+1$$

Si aplicamos (1.20), tenemos que $\text{sgn} P'_{n+1}(x_{n+1,k}) = (-1)^{n+1-k}$, por lo que para que se cumpla la desigualdad necesariamente $\text{sgn} P_n(x_{n+1,k}) = (-1)^{n+1-k}$. Esto es, $P_n(x)$ tiene que cambiar de signo un número de impar de veces en el intervalo $(x_{n+1,k}, x_{n+1,k+1})$, y como esto ocurre para todo k , necesariamente P_n tiene un único cero en cada uno de los n intervalos $(x_{n+1,k}, x_{n+1,k+1})$, $k = 1, \dots, n$. \square

Este teorema tiene una consecuencia inmediata

Corolario 1.11. *Para cada $k \geq 1$, la sucesión $\{x_{nk}\}$ es estrictamente decreciente y la sucesión $\{x_{n,n-k+1}\}$ es estrictamente creciente.*

De hecho, los límites

$$\xi_k = \lim_{n \rightarrow \infty} x_{nk}, \quad \eta_k = \lim_{n \rightarrow \infty} x_{n,n-k+1}$$

existen para todo $k = 1, \dots, n$, pudiendo tomar los valores $\pm\infty$.

Estos límites nos llevan a la siguiente definición

Definición 1.10 (Verdadero Intervalo de Ortogonalidad). El intervalo cerrado $[\xi_1, \eta_1]$ es conocido como el **Verdadero Intervalo de Ortogonalidad** de la SPO asociada al funcional \mathcal{L} .

Este intervalo recién presentado es el intervalo cerrado más pequeño que contiene todos los ceros de todos los polinomios de la SPO $\{P_n\}$. Es posible probar que el conjunto de todos los ceros de todos los P_n , llamémoslo I , es un soporte de \mathcal{L} , por lo que el verdadero intervalo de ortogonalidad es además el intervalo cerrado más pequeño que es un soporte de \mathcal{L} .

En muchos casos se presenta una familia de polinomios ortogonales mediante un funcional y un intervalo (a, b) en el cual se verifica la ortogonalidad. El conjunto I de todos los ceros es un conjunto denso en (a, b) , de forma que su adherencia, $[\xi_1, \eta_1]$, coincida con el intervalo de ortogonalidad.

Sin embargo, recordemos que mediante el teorema de Favard también podemos obtener funcionales definidos positivos a partir de una sucesión de polinomios $\{P_n\}$ y sucesiones de números reales $\{\beta_n\}, \{\gamma\}$ verificando ciertas condiciones. Dicho funcional no especifica el intervalo de ortogonalidad, el cual se puede obtener entonces mediante los ceros de los polinomios $\{P_n\}$.

TODO Demostraciones de todo y preguntar si esto está bien o entendí seriamente mal

Ejemplo 4. Como ejemplo, veamos los ceros de los polinomios de Tchebichef de tipo I, que están todos situados en el intervalo $[-1, 1]$. En la imagen 1.1 hemos representado³ gráficamente los polinomios de Tchebichef para $n = 2, 3, 4, 5$, asignando los colores rojo, verde, azul y negro respectivamente, señalando los ceros de cada polinomio con su color correspondiente.

Obsérvese que en efecto estos ceros están entrelazados, de forma que un cero rojo está entre dos ceros verdes; uno verde entre dos azules y uno azul entre dos negros.

Además, el primer cero de cada polinomio es cada vez menor, de forma que el primer cero negro es el más pequeño. Análogamente, el último cero es cada vez mayor, siendo el negro el más pequeño de todos. Este hecho ejemplifica el corolario 1.11 con $k = 1$.

Los ceros de los polinomios ortogonales tienen un papel importante, por ejemplo, en fórmulas de integración numérica, ya que tomar los nodos en los ceros de los polinomios ortogonales reduce considerablemente el error en la integración. Una aplicación bastante directa podrían ser las fórmulas de cuadratura gaussiana (véase [Chi11, Ch. I, Sección 6]). Incluso tienen una interpretación en el campo del equilibrio electrostático (véase [Ste18]).

³En el archivo `software/ceros.ipynb`, también disponible en la URL <https://github.com/JAntonioVR/Polinomios-Ortogonales/blob/main/software/ceros.ipynb>, puede encontrarse el código con el cual se han generado las imágenes, pudiendo ser ejecutado o incluso modificado para visualizar las gráficas con grados superiores a 5 u otras familias de polinomios.

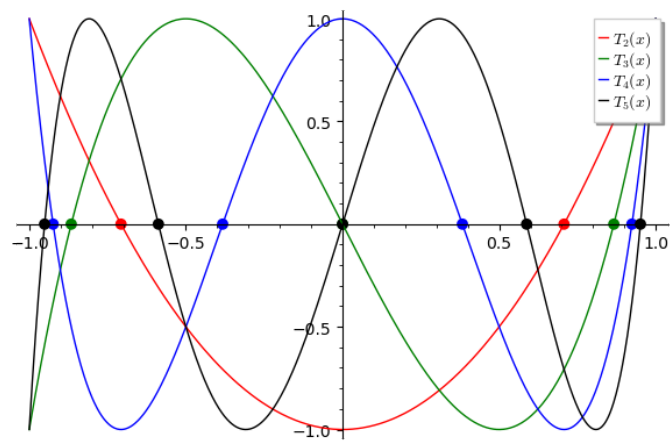


Imagen 1.1: Polinomios de Tchebichef de tipo I y sus ceros

En el capítulo 1 hemos mostrado una amplia introducción a la ortogonalidad y presentado ejemplos concretos de polinomios ortogonales. En este capítulo presentaremos las familias concretas de polinomios ortogonales más importantes: los *polinomios ortogonales clásicos*. Estos son los polinomios de Hermite, Laguerre, Jacobi y Bessel. Estas familias presentan la peculiaridad de ser las únicas que verifican ciertas propiedades, entre las que destaca la *ecuación diferencial de Pearson*. Estos polinomios, entre otras disciplinas pueden ser encontrados en problemas de Sturm-Liouville cuando se utilizan ecuaciones diferenciales hipergeométricas.

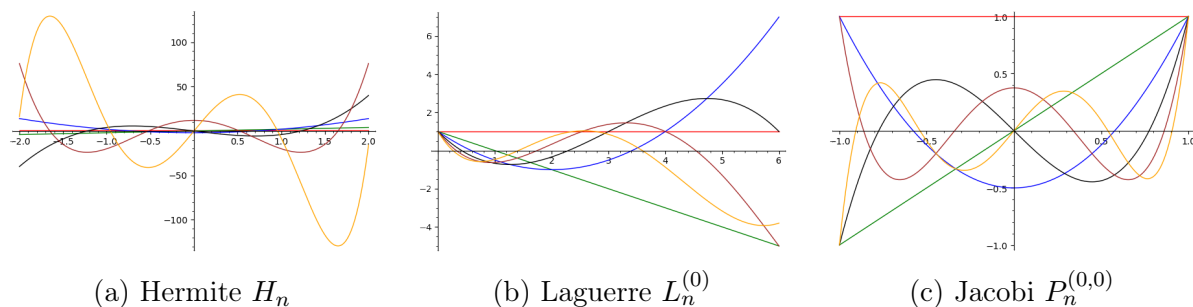


Imagen 2.1: Polinomios Ortogonales Clásicos

En las imágenes 2.1 podemos ver la representación gráfica, para parámetros concretos que definiremos próximamente, de los polinomios ortogonales de Hermite, Laguerre y Jacobi de grados $n = 0, 1, \dots, 5$. Estas tres son las familias que mayor atención recibirán de nuestra parte al ser aquellas cuya ortogonalidad se manifiesta en intervalos reales.

2.1. La ecuación de Pearson

Para empezar, introduciremos una nueva notación para los funcionales de momentos que también se suele utilizar. Sea \mathbf{u} un funcional de momentos y $\{\mu_n\}$ su sucesión de

momentos, entonces denotamos

$$\begin{aligned}\mathbf{u} : \mathbb{P} &\longrightarrow \mathbb{R} \\ x^n &\longmapsto \langle \mathbf{u}, x^n \rangle = \mu_n.\end{aligned}\tag{2.1}$$

A lo largo de este capítulo denotaremos como $\langle \mathbf{u}, \pi(x) \rangle$ a la aplicación del funcional \mathbf{u} a un polinomio $\pi(x)$.

Definimos a continuación dos operadores que actúan sobre los funcionales.

Definición 2.1 (Producto por un polinomio). Para cada polinomio π , definimos un nuevo funcional de momentos a partir de \mathbf{u} como

$$\begin{aligned}\pi\mathbf{u} : \mathbb{P} &\longrightarrow \mathbb{R} \\ \phi &\longmapsto \langle \pi\mathbf{u}, \phi \rangle = \langle \mathbf{u}, \pi\phi \rangle\end{aligned}\tag{2.2}$$

Definición 2.2 (Derivada). Definimos el funcional derivada como

$$\begin{aligned}D(\mathbf{u}) : \mathbb{P} &\longrightarrow \mathbb{R} \\ \phi &\longmapsto \langle D(\mathbf{u}), \phi \rangle = -\langle \mathbf{u}, \phi' \rangle\end{aligned}\tag{2.3}$$

A partir de estos dos operadores y esta nueva notación daremos una primera definición de lo que es una SPO clásica.

REVIEW Nona me dice que ponga ‘**una** SPO’, no ‘**la** SPO’ para un funcional. Suelo asumir la unicidad salvo constante o salvo estandarización, de ahí que tenga una sección hablando de estandarizaciones. Aún así, ¿pongo mejor ‘una’ para no asumir esa unicidad?

Definición 2.3 (SPO Clásica). La SPO $\{P_n\}$ para el funcional \mathbf{u} se dice que es **clásica** (Hermite, Laguerre, Jacobi o Bessel) si existen polinomios $\sigma(x), \tau(x)$ con $\deg(\sigma) \leq 2$ y $\deg(\tau) = 1$ tales que el funcional \mathbf{u} verifica la ecuación diferencial

$$D(\sigma\mathbf{u}) = \tau\mathbf{u}\tag{2.4}$$

La ecuación (2.4) es conocida como la **ecuación de Pearson**.

En el caso de que la SPO sea ortogonal respecto a una función peso ρ con momentos finitos, esto es, \mathbf{u} es de la forma

$$\langle \mathbf{u}, \pi(x) \rangle = \int_a^b \pi(x) \rho(x) dx,\tag{2.5}$$

donde (a, b) es cierto intervalo de la recta real donde $\rho > 0$ y $\left| \int_a^b x^k \rho(x) dx \right| < \infty$ ($k \geq 0$), podemos deducir una condición equivalente.

Lema 2.1. Sean $\rho(x)$ una función peso positiva en el intervalo (a, b) con momentos finitos, \mathbf{u} el funcional definido como en (2.5) y $\{P_n\}$ la SPO respecto a \mathbf{u} . Entonces las siguientes afirmaciones son equivalentes:

1. La función peso $\rho(x)$ es solución de la ecuación diferencial

$$[\sigma(x)\rho(x)]' = \tau(x)\rho(x).\tag{2.6}$$

2. La SPO $\{P_n\}$ es clásica, i.e. \mathbf{u} verifica la ecuación de Pearson (2.4).

Esta equivalencia se cumple siempre que se impongan sobre $\rho(x)$ las condiciones de frontera

$$\lim_{x \rightarrow a} \sigma(x)\rho(x)x^n = \lim_{x \rightarrow b} \sigma(x)\rho(x)x^n, \quad n \geq 0. \quad (2.7)$$

Demostración. \Rightarrow Queremos comprobar la igualdad de funcionales $D(\sigma\mathbf{u}) = \tau\mathbf{u}$. Para ello, tengamos en cuenta que dos funcionales lineales son iguales si, y solo si actúan igual sobre cualquier polinomio $\pi(x)$. Entonces tenemos que

$$\langle D(\sigma\mathbf{u}), \pi(x) \rangle = -\langle \sigma\mathbf{u}, \pi'(x) \rangle = -\langle \mathbf{u}, \sigma(x)\pi'(x) \rangle,$$

aplicando (2.5) e integración por partes, llegamos a

$$\begin{aligned} \langle D(\sigma\mathbf{u}), \pi(x) \rangle &= -\int_a^b \sigma(x)\pi'(x)\rho(x)dx = \left\{ \begin{array}{ll} u = \sigma(x)\rho(x) & du = [\sigma(x)\rho(x)]'dx \\ dv = \pi'(x)dx & v = \pi(x) \end{array} \right\} \\ &= -\underbrace{[\sigma(x)\rho(x)\pi(x)]_a^b}_{=0 \text{ por (2.7)}} + \int_a^b \pi(x)[\sigma(x)\rho(x)]'dx \quad (\text{aplicando (2.6)}) \\ &= \int_a^b \pi(x)\tau(x)\rho(x)dx = \langle \mathbf{u}, \tau(x)\pi(x) \rangle = \langle \tau\mathbf{u}, \pi(x) \rangle \end{aligned}$$

\Leftarrow Queremos probar que la función peso $\rho(x)$ es solución de la ecuación (2.6), sabiendo que el funcional \mathbf{u} definido como en (2.5) verifica la ecuación de Pearson (2.4) y que $\rho(x)$ cumple (2.7).

Por un lado, tenemos que para cualquier polinomio $\pi(x)$:

$$\int_a^b \pi(x)\tau(x)\rho(x)dx = \langle \tau\mathbf{u}, \pi(x) \rangle = \langle D(\sigma\mathbf{u}), \pi(x) \rangle = -\int_a^b \pi'(x)\sigma(x)\rho(x)dx.$$

Por tanto, si ahora aplicamos integración por partes

$$\begin{aligned} 0 &= \int_a^b \pi(x)\tau(x)\rho(x)dx + \int_a^b \pi'(x)\sigma(x)\rho(x)dx \\ &= \left\{ \begin{array}{ll} u = \sigma(x)\rho(x) & du = [\sigma(x)\rho(x)]'dx \\ dv = \pi'(x)dx & v = \pi(x) \end{array} \right\} \\ &= \int_a^b \pi(x)\tau(x)\rho(x)dx + \underbrace{[\sigma(x)\rho(x)\pi(x)]_a^b}_{=0 \text{ por (2.7)}} - \int_a^b \pi(x)[\sigma(x)\rho(x)]'dx \\ &= \int_a^b \pi(x)(\tau(x)\rho(x) - [\sigma(x)\rho(x)]')dx \end{aligned}$$

Si la última integral es nula para todo polinomio $\pi(x)$, necesariamente $\tau(x)\rho(x) - [\sigma(x)\rho(x)]' = 0$ y tendríamos la ecuación de Pearson (2.6). **REVIEW ¿En realidad no sería que esa ecuación se verifica para casi todo punto?**

□

Definición 2.4. Diremos que una SPO $\{P_n\}$ es **clásica respecto a una función peso** $\rho(x)$ si existen polinomios $\sigma(x), \tau(x)$ con $\deg(\sigma) \leq 2$ y $\deg(\tau) = 1$ tales que, asumiendo las condiciones de frontera (2.7), entonces $\rho(x)$ es solución de la ecuación diferencial (2.6).

Notemos que esta definición, gracias al lema 2.1, es equivalente a la definición 2.3 definiendo el funcional \mathbf{u} a partir del peso ρ como en (2.5).

2.2. Deducción de las familias de polinomios ortogonales clásicos

A partir del lema 2.1 podremos obtener las principales familias de polinomios ortogonales. Hemos impuesto que $\tau(x) = Ax + B$ sea un polinomio de grado exactamente 1, por lo que tenemos cinco posibles casos a partir de las posibilidades de $\sigma(x)$:

1. **Caso I:** σ es constante. Podemos suponer, sin pérdida de generalidad, que $\sigma(x) = 1$. En este caso obtendremos los llamados polinomios de Hermite.
2. **Caso II:** σ es de grado 1. Suponemos que $\sigma(x) = x - a$. Haciendo el cambio de variable lineal $x = -\frac{t}{A} + a$ ¹ se tiene $\sigma(x) = x$ y $\tau(x) = -x + B$. Con estos valores calcularemos los polinomios de Laguerre.
3. **Caso III:** σ es de grado 2 y tiene dos raíces reales distintas. Asumimos $\sigma(x) = (x-a)(b-x)$. Con el cambio de variable $x = \frac{b-a}{2}t + \frac{a+b}{2}$ podemos escribir $\sigma(x) = 1-x^2$ y $\tau(x) = Ax + B$. Y así deduciremos los polinomios de Jacobi.
4. **Caso IV:** σ es de grado 2 y tiene una única raíz real doble: $\sigma(x) = (x-a)^2$. En este caso se obtienen los polinomios de Bessel, que se corresponden con un funcional que no es definido positivo. Por ello, su estudio se complica considerablemente, por lo que será omitido. Más información en [Gro80].
5. **Caso V:** σ es un polinomio de grado 2 y tiene dos raíces complejas conjugadas $\sigma(x) = (x-a-bi)(x-a+bi)$. Con esta asunción la solución que se obtiene no admite momentos finitos, por lo que no puede ser considerada.

De acuerdo al lema 2.1, si resolvemos la ecuación diferencial (2.6) asumiendo las condiciones de frontera (2.7) y comprobando que los momentos resultantes son finitos, habremos hallado las funciones peso $\rho(x)$ respecto a las cuales la SPO es clásica. Nos centraremos principalmente en los casos I, II y III, en los cuales obtenemos un funcional definido positivo.

De forma genérica, podemos comprobar de manera sencilla que una fórmula explícita para $\rho(x)$ es

$$\rho(x) = \frac{1}{\sigma(x)} \exp \left\{ \int_{x_0}^x \frac{\tau(t)}{\sigma(t)} dt \right\}, \quad (2.8)$$

donde x toma valores en aquellos puntos en los que $\sigma(x) \neq 0$ y $\rho(x) > 0$.

¹Renombramos $t \equiv x$ tras el cambio de variable.

2.2.1. Caso I: Polinomios de Hermite

Supongamos que $\sigma(x) = 1$. Tenemos entonces que

$$\begin{aligned}\rho(x) &= \exp \left\{ \int_0^x (At + B) dt \right\} \\ &= \exp \left\{ \frac{A}{2} x^2 + Bx + \log(C) \right\} \\ &= C \exp \left\{ \frac{A}{2} x^2 + Bx \right\}, \quad \forall x \in \mathbb{R}\end{aligned}$$

Para que los momentos sean finitos, necesariamente $A < 0$. Por su parte, B es indiferente para la convergencia de los momentos una vez se fija A . Por este motivo, podemos elegir $A = -2, B = 0$. Tomamos $C = 1$ y llegamos a

$$\sigma(x) = 1, \quad \tau(x) = -2x, \quad \rho(x) = e^{-x^2}, \quad \forall x \in \mathbb{R}. \quad (2.9)$$

Por tanto, la sucesión de polinomios ortogonales con respecto al funcional

$$\langle \mathbf{u}, \pi(x) \rangle = \int_{-\infty}^{\infty} \pi(x) e^{-x^2} dx \quad (2.10)$$

es clásica, y sus elementos son los **polinomios de Hermite**, normalmente denotados como $\{H_n(x)\}_{n \geq 0}$.

2.2.2. Caso II: Polinomios de Laguerre

Supongamos que $\sigma(x) = x$ y $\tau(x) = -x + B$. En este caso, para que σ no se anule debemos restringirnos a intervalos en los que $x \neq 0$. Así,

$$\begin{aligned}\rho(x) &= \frac{1}{x} \exp \left\{ \int_{x_0}^x \left(-1 + \frac{B}{t} \right) dt \right\} \\ &= \frac{1}{x} \exp \{ -x + B \log |x| + \log(C) \} \\ &= C \frac{|x|^B}{x} e^{-x}.\end{aligned}$$

Si asumimos $x > 0$, entonces $\rho(x) = Cx^{B-1}e^{-x}$, resultando los momentos finitos siempre que $B > 0$. Por su parte, si $x < 0$ los momentos de $\rho(x) = C(-x)^{B-1}e^{-x}$ no resultan finitos. En resumen, podemos definir $\alpha = B - 1$ y tomar $C = 1$, de forma que

$$\sigma(x) = x, \quad \tau(x) = -x + \alpha + 1, \quad \rho(x) = x^\alpha e^{-x} \quad \forall x \in [0, +\infty), \quad \alpha > -1. \quad (2.11)$$

Así, la SPO con respecto al funcional

$$\langle \mathbf{u}, \pi(x) \rangle = \int_0^\infty \pi(x) x^\alpha e^{-x} dx, \quad \alpha > -1 \quad (2.12)$$

es clásica, y sus elementos son los **polinomios de Laguerre**, que son denotados con $\{L_n^{(\alpha)}(x)\}_{n \geq 0}$.

2.2.3. Caso III: Polinomios de Jacobi

Por último, consideremos $\sigma(x) = 1 - x^2$ y $\tau(x) = Ax + B$. Ahora σ tiene raíces simples en $x = -1, 1$. Por este motivo, la solución ρ de la ecuación de Pearson debe estar necesariamente definida en $(-\infty, -1)$, $(-1, 1)$ ó $(1, +\infty)$.

$$\begin{aligned}\rho(x) &= \frac{1}{1-x^2} \exp \left\{ \int_{x_0}^x \frac{At+B}{1-t^2} dt \right\} \\ &= \frac{1}{1-x^2} \exp \left\{ -\frac{A+B}{2} \log|1-x| - \frac{A-B}{2} \log|1+x| + \log(C) \right\} \\ &= \frac{C}{1-x^2} |1-x|^{-\frac{A+B}{2}} |1+x|^{-\frac{A-B}{2}}\end{aligned}$$

Si tomamos $x < -1$ ó $x > 1$ los momentos de la función ρ resultante no son finitos al no estar acotada en $\pm\infty$. La única posibilidad con la cual los momentos resultan finitos es tomando $x \in (-1, 1)$ y valores A, B tales que $\frac{A+B}{2} < 0$, $\frac{A-B}{2} < 0$. Por tanto, si definimos $\alpha = -\frac{A+B}{2} - 1$ y $\beta = -\frac{A-B}{2} - 1$, obtenemos

$$\begin{aligned}\sigma(x) &= 1 - x^2, & \tau(x) &= -(\alpha + \beta + 2)x + (\beta - \alpha), \\ \rho(x) &= (1-x)^\alpha (1+x)^\beta \quad \forall x \in [-1, 1], \quad \alpha, \beta > -1.\end{aligned}\tag{2.13}$$

Así, la SPO con respecto al funcional

$$\langle \mathbf{u}, \pi(x) \rangle = \int_{-1}^1 \pi(x) (1-x)^\alpha (1+x)^\beta dx, \quad \alpha, \beta > -1 \tag{2.14}$$

es clásica, y sus elementos son los **polinomios de Jacobi**, denotados como $\{P_n^{(\alpha, \beta)}(x)\}_{n \geq 0}$.

Resumiendo, tenemos tres familias de polinomios ortogonales clásicos además de la Bessel: Hermite, Laguerre y Jacobi. En la tabla 2.1 recogemos los datos más importantes sobre cada una de ellas: la notación, el intervalo de ortogonalidad y las funciones $\sigma(x)$, $\tau(x)$ y $\rho(x)$ involucradas en las ecuaciones de Pearson.

Familia	Notación	Intervalo	$\sigma(x)$	$\tau(x)$	$\rho(x)$
Hermite	$H_n(x)$	$(-\infty, \infty)$	1	$-2x$	e^{-x^2}
Laguerre	$L_n^{(\alpha)}(x)$	$[0, \infty)$	x	$-x + \alpha + 1$	$x^\alpha e^{-x}$
Jacobi	$P_n^{(\alpha, \beta)}(x)$	$[-1, 1]$	$1 - x^2$	$-(\alpha + \beta + 2)x + (\beta - \alpha)$	$(1-x)^\alpha (1+x)^\beta$

Tabla 2.1: Resumen sobre las SPO clásicas

2.3. Caracterizaciones

Hasta el momento hemos podido conocer las distintas familias de polinomios ortogonales clásicos y su peculiaridad respecto a las distintas formas de presentar la ecuación de Pearson. Sin embargo, estos polinomios cumplen varias propiedades destacables que además los caracterizan como clásicos. Es decir, son los únicos polinomios ortogonales que las verifican. En esta sección estudiaremos algunas de estas propiedades.

2.3.1. Ortogonalidad de las derivadas

La primera de estas caracterizaciones atiende a la sucesión $\{P'_n\}$ formada por las derivadas de los elementos de una SPO $\{P_n\}$, que en el caso de los polinomios ortogonales clásicos resulta no sólo ser también una SPO, sino que además también es clásica.

Teorema 2.1. *Sea $\{P_n\}$ una SPO clásica respecto a la función peso $\rho(x)$. Entonces la sucesión $\{P'_n\}$ es una SPO respecto a $\sigma(x)\rho(x)$.*

Demostración. Al ser $\{P_n\}$ una SPO y $\tau(x)$ un polinomio de grado 1, tenemos que para $k < n$:

$$\begin{aligned} 0 &= \int_a^b P_n(x)x^{k-1}\tau(x)\rho(x)dx \\ &= \int_a^b P_n(x)x^{k-1}[\sigma(x)\rho(x)]'dx \quad (\text{por (2.6).}) \end{aligned}$$

Si aplicamos integración por partes tomando

$$\left\{ \begin{array}{ll} u = P_n(x)x^{k-1} & du = P'_n(x)x^{k-1} + (k-1)P_n(x)x^{k-2}dx \\ dv = [\sigma(x)\rho(x)]'dx & v = \sigma(x)\rho(x) \end{array} \right\},$$

tenemos

$$0 = \underbrace{\left[P_n(x)x^{k-1}\sigma(x)\rho(x) \right]_a^b}_{=0 \text{ por (2.7)}} - \int_a^b P'_n(x)x^{k-1}\sigma(x)\rho(x)dx - (k-1) \underbrace{\int_a^b P_n(x)x^{k-2}\sigma(x)\rho(x)dx}_{=0 \text{ por la ortogonalidad de } \{P_n\}}.$$

Por tanto $\int_a^b P'_n(x)x^{k-1}[\sigma(x)\rho(x)]dx = 0$ para $k < n$. Comprobemos ahora que $\int_a^b P'_n(x)x^{n-1}[\sigma(x)\rho(x)]dx \neq 0$.

Si procedemos análogamente para $k = n$ llegamos a

$$\begin{aligned} \int_a^b P_n(x)x^{n-1}\tau(x)\rho(x)dx &= - \int_a^b P'_n(x)x^{n-1}\sigma(x)\rho(x)dx \\ &\quad - (n-1) \int_a^b P_n(x)x^{n-2}\sigma(x)\rho(x)dx. \end{aligned}$$

De esta igualdad deducimos

$$\int_a^b P'_n(x)x^{n-1}\sigma(x)\rho(x)dx = - \int_a^b P_n(x)[x^{n-1}\tau(x) + (n-1)x^{n-2}\sigma(x)]\rho(x)dx.$$

Y esta integral no es cero únicamente si $x^{n-1}\tau(x) + (n-1)x^{n-2}\sigma(x)$ tiene grado n , o equivalentemente si $\Delta(x) := x\tau(x) + (n-1)\sigma(x)$ tiene grado 2. Si $\sigma(x)$ tiene grado 0 o 1 es claro que $\Delta(x)$ tendrá grado 2. Si $\sigma(x)$ tiene grado 2 entonces estamos en el caso III, en el que $\sigma(x) = 1 - x^2$ y $\tau(x) = -(\alpha + \beta + 2)x + \beta - \alpha$. Con esta asunción, $x\tau(x)$ tiene coeficiente líder negativo y $(n-1)\sigma(x)$ también, por lo que el coeficiente de grado 2 de $\Delta(x)$ será no nulo y por tanto $\deg(\Delta) = 2$. Es decir, $\int_a^b P'_n(x)x^{n-1}\sigma(x)\rho(x)dx \neq 0$.

Esto es, $\{P'_n\}$ es una SPO respecto a $\sigma(x)\rho(x)$. \square

A partir de este último resultado ya podemos probar la consecuencia que ya adelantábamos: la sucesión de las derivadas también es clásica.

Corolario 2.2. *Si la sucesión $\{P_n\}$ una SPO clásica respecto a la función peso $\rho(x)$, entonces la sucesión $\{P'_n\}$ también es clásica respecto $\rho_1(x) = \sigma(x)\rho(x)$, que es solución de la ecuación de Pearson*

$$[\sigma(x)\rho_1(x)]' = \tau_1(x)\rho_1(x), \quad (2.15)$$

donde $\tau_1(x) = \tau(x) + \sigma'(x)$.

Demostración. Tenemos que $\{P'_n\}$ es una SPO respecto a $\rho_1(x) = \sigma(x)\rho(x)$. Tenemos además que

$$\begin{aligned} [\sigma(x)\rho_1(x)]' &= \sigma(x)[\sigma(x)\rho(x)]' + \sigma'(x)[\sigma(x)\rho(x)] \\ &= \sigma(x)\tau(x)\rho(x) + \sigma'(x)[\sigma(x)\rho(x)] \quad (\text{por (2.6) para } \rho(x)) \\ &= \underbrace{(\tau(x) + \sigma'(x))}_{\tau_1(x)} \rho_1(x). \end{aligned}$$

Nótese que $\tau_1(x) = \tau(x) + \sigma'(x)$ es de nuevo un polinomio de grado 1. Por tanto, ρ_1 es solución de la ecuación de Pearson, por lo que $\{P'_n\}$ es una SPO clásica. \square

Por inducción en las sucesivas derivadas $\{P_n^{(k)}\}$ podemos deducir el siguiente resultado:

Corolario 2.3. *Si la sucesión $\{P_n\}$ una SPO clásica entonces la sucesión $\{P_n^{(k)}\}$ es una SPO clásica respecto a la función peso $\rho_k(x) = \sigma^k(x)\rho(x)$, que es solución de la ecuación*

$$[\sigma(x)\rho_k(x)]' = \tau_k(x)\rho_k(x), \quad (2.16)$$

donde $\tau_k(x) = \tau(x) + k\sigma'(x)$.

2.3.2. Ecuación diferencial hipergeométrica

A partir de la ortogonalidad de las derivadas podemos encontrar una interesante ecuación diferencial que veremos también caracteriza a las familias clásicas de polinomios ortogonales.

Teorema 2.4. *Sea $\{P_n\}$ una SPO clásica respecto a la función peso $\rho(x)$. Entonces cada polinomio $P_n(x)$, $n \in \mathbb{N}_0$ es solución de la ecuación diferencial*

$$\sigma(x)P_n''(x) + \tau(x)P_n'(x) + \lambda_n P_n(x) = 0, \quad (2.17)$$

con $\lambda_n = -n \left(\tau' + (n-1) \frac{\sigma''}{2} \right) \in \mathbb{R}$.

Demostración. Al ser $\{P_n\}$ una SPO clásica, por el corolario 2.2 $\{P'_n\}$ es ortogonal y clásica respecto a $\rho_1(x) = \sigma(x)\rho(x)$. Esto es, para $k < n$:

$$0 = \int_a^b P'_n(x)(x^k)' \rho_1(x) dx. \quad (2.18)$$

Si aplicamos integración por partes y las condiciones de frontera (2.7), tenemos

$$\begin{aligned}
0 &= \int_a^b x^k (P'_n(x) \rho_1(x))' dx \\
&= \int_a^b x^k (P''_n(x) \rho_1(x) + P'_n(x) \rho'_1(x)) dx \\
&= \int_a^b x^k (P''_n(x) \sigma(x) \rho(x) + P'_n(x) \tau(x) \rho(x)) dx \quad (\text{por (2.6)}) \\
&= \int_a^b x^k (P''_n(x) \sigma(x) + P'_n(x) \tau(x)) \rho(x) dx
\end{aligned}$$

para todo $k < n$. Análogamente, teniendo en cuenta que para $k = n$ la integral (2.18) no se anula, podemos comprobar que

$$\int_a^b x^n (P''_n(x) \sigma(x) + P'_n(x) \tau(x)) \rho(x) dx \neq 0.$$

Por el teorema 1.1, esto significa que la sucesión de polinomios $\{P''_n(x) \sigma(x) + P'_n(x) \tau(x)\}$ es ortogonal respecto a la función peso ρ . Pero por el corolario 1.3, la SPO respecto a ρ es única salvo constante multiplicativa, de forma que para todo $n \in \mathbb{N}_0$ existe una constante $\lambda_n \in \mathbb{R}$ tal que

$$P''_n(x) \sigma(x) + P'_n(x) \tau(x) = -\lambda_n P_n(x).$$

Esto es, $P_n(x)$ es solución de (2.17).

Comprobemos ahora la expresión de λ_n igualando las potencias x^n en (2.17). Si consideramos $\sigma(x)$ un polinomio de grado 2 (considerando que $\mathbb{P}_0 \subset \mathbb{P}_1 \subset \mathbb{P}_2$), tendríamos que su coeficiente líder es $\sigma''(x)/2$. Análogamente, el coeficiente líder de $\tau(x)$ es τ' . Denotamos $\sigma'' \equiv \sigma''(x)$, $\tau' \equiv \tau'(x)$ ya que ambas son constantes. Por otro lado, si $c_n \neq 0$ es el coeficiente líder de $P_n(x)$, igualando los términos líderes de (2.17) obtenemos

$$\frac{\sigma''}{2} n(n-1) c_n + \tau' n c_n + \lambda_n c_n = 0 \Leftrightarrow \lambda_n = -n \left(\tau' + (n-1) \frac{\sigma''}{2} \right)$$

□

Naturalmente, al ser la sucesión de derivadas sucesivas $\{P_n^{(k)}\}$ una SPO clásica, también verificarán su propia ecuación diferencial.

Corolario 2.5. *Si la sucesión $\{P_n\}$ una SPO clásica, entonces cada polinomio $P_n^{(k)}(x)$, $k \geq 0, n \in \mathbb{N}_0$ es solución de la ecuación diferencial*

$$\sigma(x) [P_n^{(k)}(x)]'' + \tau_k(x) [P_n^{(k)}(x)]' + \lambda_n P_n^{(k)}(x) = 0, \quad (2.19)$$

con $\lambda_n \in \mathbb{R}$ y donde $\tau_k(x) = \tau(x) + k\sigma'(x)$.

De forma general, la ecuación (2.17) es normalmente conocida como la **ecuación diferencial hipergeométrica**:

$$\sigma(x) y'' + \tau(x) y' + \lambda y = 0, \quad (2.20)$$

donde $\deg(\sigma) \leq 2$, $\deg(\tau) = 1$.

Esta ecuación veremos próximamente que es de hecho equivalente a la ecuación de Pearson.

Supongamos ahora que la ecuación (2.20) admite una solución polinómica de grado n , digamos $y_n(x) = \sum_{k=0}^n c_k x^k$, con coeficiente líder $c_n \neq 0$. Por otro lado, sea $\rho(x)$ una solución de la ecuación de Pearson (2.6):

$$[\sigma(x)\rho(x)]' = \tau(x)\rho(x). \quad (2.6)$$

Entonces podemos escribir la ecuación (2.20) en su forma **simétrica o conjugada**:

$$[\rho(x)\sigma(x)y']' + \lambda\rho(x)y = 0. \quad (2.21)$$

REVIEW Esto al final no lo uso para nada.

Como vemos, las soluciones de la ecuación de Pearson y las soluciones de la ecuación hipergeométrica están íntimamente relacionadas. Benjamin Olinde Rodrigues (1795–1851) demostró que las soluciones polinómicas de la EDO (2.20) pueden escribirse de una peculiar y compacta forma hoy llamada fórmula de Rodrigues, que veremos que también caracteriza a los polinomios ortogonales clásicos.

2.3.3. Fórmula de Rodrigues

Originalmente, Olinde Rodrigues (1816), James Ivory (1824) y Carl Gustav Jacobi (1827) introdujeron una fórmula inicialmente definida para los polinomios de Legendre, posteriormente extendida a todas las soluciones polinómicas de la ecuación hipergeométrica. El nombre de fórmula de Rodrigues fue propuesto por Eduard Heine en 1878, después de que Hermite recuperara en 1860 el trabajo original de Rodrigues y señalara que fue él el primero en descubrir esta fórmula. Ya anunciada la peculiaridad de esta fórmula, presentamos su aspecto y demostración en el siguiente resultado.

Teorema 2.6 (Fórmula de Rodrigues). *Las soluciones polinómicas de la ecuación hipergeométrica (2.20) se pueden expresar como*

$$y_n(x) = \frac{1}{\rho(x)} \frac{d^n}{dx^n} (\rho(x)\sigma^n(x)), \quad (2.22)$$

donde $\rho(x)$ es una solución de la ecuación de Pearson (2.6).

Demostración. Por simplicidad en la notación, en esta demostración será omitida y asumida la dependencia de x en las distintas funciones. Al ser ρ una solución de la ecuación de Pearson (2.6), tenemos que

$$\begin{aligned} \sigma(\rho\sigma^n)' &= \sigma(\sigma^{n-1} \cdot \sigma\rho)' \\ &= \sigma((n-1)\sigma^{n-2}\sigma'\sigma\rho + \sigma^{n-1}\tau\rho) \\ &= \rho\sigma^n((n-1)\sigma' + \tau), \\ \sigma(\rho\sigma^n)' &= \rho\sigma^n((n-1)\sigma' + \tau) \end{aligned} \quad (2.23)$$

Derivamos ahora la ecuación (2.23) $n+1$ veces y dividimos entre la función peso $\rho(x)$. Para derivar la ecuación (2.23) $n+1$ veces aplicamos la fórmula de Leibniz² y tenemos en cuenta que, como σ tiene grado a lo sumo 2 el miembro de la izquierda se anularía a partir de la tercera derivada. Análogamente, como σ', τ tienen grado a lo sumo 1 el miembro derecho se anularía a partir de la segunda derivada. Así:

$$\begin{aligned} \frac{d^{n+1}}{dx^{n+1}} (\sigma(\rho\sigma^n)') &= \sigma \frac{d^{n+2}}{dx^{n+2}} (\rho\sigma^n) + (n+1)\sigma' \frac{d^{n+1}}{dx^{n+1}} (\rho\sigma^n) + \frac{(n+1)n}{2}\sigma'' \frac{d^n}{dx^n} (\rho\sigma^n), \\ \frac{d^{n+1}}{dx^{n+1}} (\rho\sigma^n((n-1)\sigma' + \tau)) &= ((n-1)\sigma' + \tau) \frac{d^{n+1}}{dx^{n+1}} (\rho\sigma^n) + (n+1)((n-1)\sigma'' + \tau') \frac{d^n}{dx^n} (\rho\sigma^n) \end{aligned}$$

Igualando los segundos miembros de ambas igualdades y dividiendo entre ρ obtenemos

$$\begin{aligned} \frac{\sigma}{\rho} \frac{d^{n+2}}{dx^{n+2}} (\rho\sigma^n) + \frac{(n+1)\sigma'}{\rho} \frac{d^{n+1}}{dx^{n+1}} (\rho\sigma^n) + \frac{(n+1)n\sigma''}{2\rho} \frac{d^n}{dx^n} (\rho\sigma^n) = \\ \frac{(n-1)\sigma' + \tau}{\rho} \frac{d^{n+1}}{dx^{n+1}} (\rho\sigma^n) + \frac{(n+1)((n-1)\sigma'' + \tau')}{\rho} \frac{d^n}{dx^n} (\rho\sigma^n) \end{aligned} \quad (2.24)$$

Nuestro objetivo ahora es manipular (2.24) de forma que podamos comprobar que una función y_n dada como en (2.22) es una solución de la ecuación hipergeométrica (2.20). Si en (2.24) agrupamos términos y teniendo en cuenta la definición de y_n , tenemos la ecuación

$$\frac{\sigma}{\rho} \frac{d^{n+2}}{dx^{n+2}} (\rho\sigma^n) + \frac{2\sigma' - \tau}{\rho} \frac{d^{n+1}}{dx^{n+1}} (\rho\sigma^n) - \left(\frac{n^2 - n - 2}{2}\sigma'' + (n+1)\tau' \right) y_n = 0. \quad (2.25)$$

Seguidamente, debemos ajustar la ecuación de manera que los factores $1/\rho$ solo tengan n derivadas a su derecha. Para ello primero nos ayudamos de la siguiente identidad:

$$\frac{\sigma}{\rho} \frac{d^{n+2}}{dx^{n+2}} (\rho\sigma^n) = \sigma \left(\frac{1}{\rho} \frac{d^n}{dx^n} (\rho\sigma^n) \right)'' - 2\sigma(1/\rho)' \frac{d^{n+1}}{dx^{n+1}} (\rho\sigma^n) - \sigma(1/\rho)'' \frac{d^n}{dx^n} (\rho\sigma^n). \quad (2.26)$$

Esta igualdad se puede deducir sin más que derivar dos veces el producto $\left(\frac{1}{\rho}\right) \cdot \left(\frac{d^n}{dx^n} (\rho\sigma^n)\right)$ utilizando la fórmula de Leibniz e intercambiando miembros en la igualdad. Si calculamos explícitamente las derivadas primera y segunda de $1/\rho$ y aplicamos que, como consecuencia de la ecuación de Pearson (2.6), podemos deducir la identidad

$$\rho' = \rho \frac{\tau - \sigma'}{\sigma}. \quad (2.27)$$

Entonces la ecuación (2.26) puede reducirse a

$$\frac{\sigma}{\rho} \frac{d^{n+2}}{dx^{n+2}} (\rho\sigma^n) = \sigma y_n'' + \frac{2(\tau - \sigma')}{\rho} \frac{d^{n+1}}{dx^{n+1}} (\rho\sigma^n) - \left(\sigma'' - \tau' - \frac{\tau - \sigma'}{\rho} \right) y_n. \quad (2.28)$$

Si ahora sustituimos (2.28) en (2.25), con algunas simplificaciones obtenemos

$$\sigma y_n'' + \frac{\tau}{\rho} \frac{d^{n+1}}{dx^{n+1}} (\rho\sigma^n) - \left(\frac{n^2 - n}{2}\sigma'' + n\tau' - \frac{\tau(\tau - \sigma')}{\sigma} \right) y_n = 0. \quad (2.29)$$

²Fórmula de Leibniz: Sean $u, v : I \rightarrow \mathbb{R}$ dos funciones reales n -veces derivables en $I \subseteq \mathbb{R}$, entonces se verifica en I que $(uv)^{(n)} = \sum_{k=0}^n \binom{n}{k} u^{(k)} v^{(n-k)}$.

El último paso es aplicar la identidad

$$\frac{\tau}{\rho} \frac{d^{n+1}}{dx^{n+1}} (\rho \sigma^n) = \tau y'_n - \frac{\tau(\tau - \sigma')}{\sigma} y_n. \quad (2.30)$$

De forma similar a la ecuación (2.28), esta expresión se obtiene al derivar una sola vez el producto $\left(\frac{1}{\rho}\right) \cdot \left(\frac{d^n}{dx^n}(\rho \sigma^n)\right)$, intercambiando miembros de la igualdad y teniendo en cuenta el valor de ρ' deducido de la ecuación de Pearson.

Sustituyendo (2.30) en (2.28) llegamos finalmente a

$$\sigma y''_n + \tau y'_n - \left(\frac{n^2 - n}{2} \sigma'' + n \tau'\right) y_n = 0. \quad (2.31)$$

Como $\deg \sigma \leq 2, \deg \tau \leq 1$, tenemos que $\frac{n^2 - n}{2} \sigma'' + n \tau'$ es una constante, por lo que concluimos que y_n en efecto es una solución de (2.20).

Para finalizar, tenemos que probar que la solución y_n dada por la fórmula de Rodrigues (2.22) es un polinomio de grado n . Observemos que, aplicando de nuevo la fórmula de Leibniz para la derivada, tendríamos

$$y_n = \frac{1}{\rho} \sum_{k=0}^n \binom{n}{k} \frac{d^k \rho}{dx^k} \frac{d^{n-k} \sigma^n}{dx^{n-k}}. \quad (2.32)$$

Mediante inducción y utilizando (2.27) podemos probar que

$$\frac{d^k \rho}{dx^k} = \frac{\rho}{\sigma^k} P_k, \quad k \geq 0$$

donde P_k es un polinomio de grado a lo sumo k .

También podemos probar por inducción que

$$\frac{d^s \sigma^n}{dx^s} = \sigma^{n-s} Q_s, \quad 1 \leq s \leq n,$$

siendo Q_s un polinomio de grado a lo sumo s .

Teniendo en cuenta estas dos igualdades, tomando $s = n - k$ y sustituyendo en (2.32), llegamos a que

$$\begin{aligned} \frac{1}{\rho} \frac{d^n}{dx^n} (\rho \sigma^n) &= \frac{1}{\rho} \sum_{k=0}^n \binom{n}{k} \frac{\rho}{\sigma^k} P_k \sigma^k Q_{n-k} \\ &= \frac{1}{\rho} \sum_{k=0}^n \binom{n}{k} \rho P_k Q_{n-k} \\ &= \sum_{k=0}^n \binom{n}{k} P_k Q_{n-k}, \end{aligned} \quad (2.33)$$

que en definitiva es un polinomio de grado a lo sumo n .

□

Ya hemos comprobado que las soluciones polinómicas de la ecuación hipergeométrica se pueden expresar mediante la fórmula de Rodrigues. Sin embargo, aunque de momento

sabemos que, a pesar de su aspecto, y_n es un polinomio de grado a lo sumo n , podemos dar un paso más y afirmar que es de hecho un polinomio de grado exacto n . Para ello, introduciremos la ortogonalidad.

[REVIEW Algo que no sea correcto?](#)

Proposición 2.1. Las soluciones polinómicas de la ecuación diferencial hipergeométrica (2.20) y_n , expresadas mediante la fórmula de Rodrigues (2.22), son polinomios de grado exacto n .

Demostración. Recordemos que ρ es una solución de la ecuación de Pearson (2.6) positiva en un intervalo $[a, b] \subseteq \mathbb{R}$. Podemos entonces construir a partir de ρ un funcional \mathbf{u} definido como en (2.5) y su correspondiente producto escalar. Por tanto, tenemos el producto escalar

$$\langle p(x), q(x) \rangle = \langle \mathbf{u}, p(x)q(x) \rangle = \int_a^b p(x)q(x)\rho(x)dx, \quad \forall p(x), q(x) \in \mathbb{P}.$$

Sea $\pi(x) \in \mathbb{P}_n$ un polinomio de grado a lo sumo n . Calculemos $\langle \pi, y_n \rangle$.

$$\begin{aligned} \langle \pi, y_n \rangle &= \langle \mathbf{u}, \pi(x)y_n(x) \rangle = \int_a^b \pi(x)y_n(x)\rho(x)dx \\ &= \int_a^b \pi(x) \left[\frac{1}{\rho(x)} \frac{d^n}{dx^n} (\rho(x)\sigma^n(x)) \right] \rho(x)dx \\ &= \int_a^b \pi(x) \left[\frac{d^n}{dx^n} (\rho(x)\sigma^n(x)) \right] dx \\ &= \left\{ \begin{array}{l} u = \pi(x) \\ dv = \left[\frac{d^n}{dx^n} (\rho(x)\sigma^n(x)) \right] dx \end{array} \quad \begin{array}{l} du = \pi'(x)dx \\ v = \left[\frac{d^{n-1}}{dx^{n-1}} (\rho(x)\sigma^n(x)) \right] \end{array} \right\} \\ &= \underbrace{\left[\pi(x) \frac{d^{n-1}}{dx^{n-1}} (\rho(x)\sigma^n(x)) \right]_a^b}_{(1)} - \int_a^b \pi'(x) \left[\frac{d^{n-1}}{dx^{n-1}} (\rho(x)\sigma^n(x)) \right] dx \end{aligned}$$

Por ([TODO Encontrar referencia del lema que no sea el TFG de un buen chaval](#)), tenemos que

$$\frac{d^{n-1}}{dx^{n-1}} = \tilde{P}_{n-1}(x)\sigma^{n-1}(x)\rho(x),$$

siendo $\tilde{P}_{n-1}(x)$ un polinomio de grado a lo sumo $n-1$. En definitiva, la expresión (1) se puede reducir a $\psi(x)\sigma(x)\rho(x)$ siendo $\psi(x) = \sum_{k=0}^{\deg(\psi)} c_k x^k$ un polinomio. Entonces, aplicando las condiciones de frontera (2.7), tenemos que (1) vale 0, por lo que

$$\langle \pi, y_n \rangle = \int_a^b \pi(x) \left[\frac{d^n}{dx^n} (\rho(x)\sigma^n(x)) \right] dx = - \int_a^b \pi'(x) \left[\frac{d^{n-1}}{dx^{n-1}} (\rho(x)\sigma^n(x)) \right] dx.$$

Si aplicamos este procedimiento n veces, finalmente llegamos a

$$\langle \pi, y_n \rangle = (-1)^n \int_a^b \left[\frac{d^n}{dx^n} \pi(x) \right] \sigma^n(x)\rho(x)dx.$$

Entonces:

- Si $\deg(\pi) < n$, entonces su derivada n -ésima es nula y $\langle \pi, y_n \rangle = 0$.
- Si $\deg(\pi) = n$, entonces su derivada n -ésima es una constante no nula y $\langle \pi, y_n \rangle = K \neq 0$.

Esto es, y_n es un polinomio de grado menor o igual que n ortogonal a todos los polinomios de grado menor que n . Por tanto, necesariamente y_n es un polinomio de grado exacto n . \square

Recordemos ahora que los polinomios ortogonales clásicos son, en particular, soluciones polinómicas de la ecuación diferencial hipergeométrica. De acuerdo con estos dos últimos resultados, podemos escribir los polinomios ortogonales clásicos mediante la fórmula de Rodrigues (2.22). Lo habitual es multiplicar la fórmula de Rodrigues por una constante B_n , la cual depende de la familia y de la estandarización que se desee. De esta forma, si $\{P_n\}$ es una SPO clásica respecto a la función peso $\rho(x)$, entonces

$$P_n(x) = \frac{B_n}{\rho(x)} \frac{d^n}{dx^n} (\rho(x) \sigma^n(x)). \quad (2.34)$$

Y como ya anunciamos anteriormente, la fórmula de Rodrigues caracteriza a los polinomios ortogonales clásicos.

Teorema 2.7. *Si una SPO $\{P_n\}$ respecto a una función peso $\rho(x)$ se puede expresar mediante la fórmula de Rodrigues (2.34), entonces $\{P_n\}$ es una SPO clásica.*

Demostración. Si tomamos $n = 1$ en (2.34), obtenemos

$$\frac{P_1(x)}{B_1} \rho(x) = [\rho(x) \sigma(x)]'.$$

Si denotamos $\tau(x) := \frac{1}{B_1} P_1(x)$, que es claramente un polinomio de grado 1, tenemos que $\rho(x)$ verifica la ecuación de Pearson (2.6), por lo que $\{P_n\}$ es clásica. \square

El conjunto completo de teoremas e implicaciones demostradas hasta el momento conforma el siguiente teorema de caracterización.

Teorema 2.8. *Sea $\{P_n\}$ una SPO reales respecto a un funcional \mathbf{u} definido como en (2.5) a partir de la función peso $\rho(x)$. Sean $\sigma(x), \tau(x)$ dos polinomios de grado a lo sumo 2 y exactamente 1 respectivamente. Entonces las siguientes afirmaciones son equivalentes:*

1. $\{P_n\}$ es una SPO clásica respecto a la función peso $\rho(x)$, solución de la EDO (2.6).
2. La sucesión de derivadas $\{P'_n\}$ es una SPO clásica respecto a la función peso $\rho_1(x) = \sigma(x)\rho(x)$, solución de la EDO (2.15).
3. La sucesión de derivadas k -ésimas $\{P_n^{(k)}\}$ es una SPO clásica respecto a la función peso $\rho_k(x) = \sigma^k(x)\rho(x)$, solución de la EDO (2.16).
4. Cada polinomio $P_n(x)$ es solución de la ecuación diferencial hipergeométrica (2.20).
5. $\{P_n\}$ puede ser expresado mediante la fórmula de Rodrigues (2.34).

Demostración. Revisamos las implicaciones ya probadas y aclararemos las que aún no han sido demostradas.

1. \Rightarrow 2. Es el teorema 2.1.

2. \Rightarrow 1. Supongamos que la sucesión $\{P'_n\}$ es clásica respecto a la función peso $\rho_1(x) = \sigma(x)\rho(x)$, solución de $[\sigma(x)\rho_1(x)]' = \tau_1(x)\rho(x)$ con $\tau_1(x) = \tau(x) + \sigma'(x)$. Si derivamos el primer miembro y sustituimos la expresión de ρ_1 y cancelamos $\sigma(x)$, obtenemos

$$2\sigma'(x)\rho(x) + \sigma(x)\rho'(x) = \tau_1(x)\rho(x) \Leftrightarrow$$

$$\sigma'(x)\rho(x) + \sigma(x)\rho'(x) = \tau_1(x)\rho(x) - \sigma'(x)\rho(x) \Leftrightarrow$$

$$[\sigma(x)\rho(x)]' = \tau(x)\rho(x),$$

que es justamente la ecuación de Pearson (2.6).

2. \Leftrightarrow 3. Es trivial sin más que aplicar sucesivamente la equivalencia 1. \Leftrightarrow 2.

1. \Rightarrow 4. Es el teorema 2.4.

4. \Rightarrow 5. Es el teorema 2.6 junto con la proposición 2.1.

5. \Rightarrow 1. Es el teorema 2.7.

□

REVIEW Corregido, pero un detalle. Resulta que, aunque tenemos que $1 \Rightarrow 4$ porque el teorema 2.4 dice que si $\{P_n\}$ es clásica entonces es solución de la ecuación hipergeométrica, si nos fijamos en la demostración lo primero que hago es usar que las derivadas son ortogonales, por tanto, ¿podría decir que (4) es consecuencia de (3)? De poder hacerlo podría encadenar $1 \Rightarrow 2 \Rightarrow 3 \Rightarrow 4 \Rightarrow 5 \Rightarrow 1$.

Ejemplo 5 (Fórmula de Rodrigues de los polinomios de Jacobi). Pongamos en práctica lo aprendido a partir de un ejemplo en el que emplearemos lo que sabemos hasta el momento de los polinomios de Jacobi. Atendiendo a la tabla 2.1, sabemos que

$$\sigma(x) = 1 - x^2, \quad \tau(x) = -(\alpha + \beta + 2)x + (\beta - \alpha) \quad \rho(x) = (1 - x)^\alpha(1 + x)^\beta,$$

con $x \in [-1, 1]$, $\alpha, \beta > -1$. Con estos datos, tenemos que los polinomios de Jacobi $P_n^{(\alpha, \beta)}(x)$ son soluciones de la ecuación hipergeométrica

$$(1 - x^2)y'' - ((\alpha + \beta + 2)x + \alpha - \beta)y' + n(n + \alpha + \beta + 1)y = 0.$$

Y la fórmula de Rodrigues, teniendo en cuenta que se suele tomar $B_n = \frac{(-1)^n}{2^n n!}$, sería:

$$P_n^{(\alpha, \beta)}(x) = \frac{(-1)^n}{2^n n!} (1 - x)^{-\alpha} (1 + x)^{-\beta} \frac{d^n}{dx^n} (1 - x)^{n+\alpha} (1 + x)^{n+\beta}.$$

n	$H_n(x)$	RR3T	Rodrigues	Nativa
0	1	1.1897×10^{-5}	0.0086	0.00019
1	$2x$	0.0133	0.0104	0.00028
2	$4x^2 - 2$	0.0449	0.0136	0.00035
3	$8x^3 - 12x$	0.1073	0.0133	0.00037
4	$16x^4 - 48x^2 + 12$	0.2161	0.0162	0.00041
5	$32x^5 - 160x^3 + 120x$	0.4162	0.0182	0.00051
6	$64x^6 - 480x^4 + 720x^2 - 120$	0.6752	0.0186	0.0186
7	$128x^7 - 1344x^5 + 3360x^3 - 1680x$	1.2009	0.0196	0.00054
8	$256x^8 - 3584x^6 + 13440x^4 - 13440x^2 + 1680$	2.0665	0.0305	0.00061
9	$512x^9 - 9216x^7 + 48384x^5 - 80640x^3 + 30240x$	3.5725	0.0229	0.00064

Tabla 2.2: Polinomios de Hermite y tiempos de cálculo de distintos métodos

Ejemplo 6 (Cálculo de la SPO de Hermite mediante la fórmula de Rodrigues). Recordemos el ejemplo 3, en el cual utilizamos la versión más primitiva de la RR3T para calcular de forma explícita los polinomios mónicos de Hermite. La fórmula de Rodrigues nos proporciona un nuevo método de cálculo explícito de polinomios, que es el que pondremos en práctica en este caso. En el archivo `software/Rodrigues.ipynb` se ha implementado el código necesario para este ejemplo.

De manera análoga a lo realizado en el ejemplo 5 y sabiendo que, por convenio, lo usual en los polinomios de Hermite es tomar $B_n = (-1)^n$, podemos deducir la siguiente fórmula de Rodrigues.

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}$$

Así, obtenemos los polinomios de Hermite estándar, los cuales se pueden consultar en la segunda columna de la tabla 2.2. En este caso se ha hecho también un análisis de tiempos de ejecución siguiendo la misma metodología que en el ejemplo 3. Es decir, se han realizado 10 ejecuciones independientes de cada polinomio y el dato presentado en la cuarta columna de la tabla es la media de estas diez ejecuciones. Como puede observarse, el cálculo mediante la fórmula de Rodrigues mejora considerablemente los tiempos de ejecución de la RR3T, de forma que para ninguno de estos grados alcanza media décima de ejecución, mientras que la RR3T alcanzaba el orden de segundos a partir de grado 7.

De hecho, con la fórmula de Rodrigues se ha conseguido calcular el polinomio de Hermite de grado 500 en tan sólo 15 segundos, lo cual contrasta con los casi 60 segundos que tardó la RR3T en calcular el polinomio de grado 15.

Finalmente, y para cerrar esta sección, mencionaremos que en el apéndice A puede encontrarse una serie de tablas con toda la información descrita para cada familia de polinomios clásicos.

2.4. Polinomios de Hermite, Laguerre y Jacobi

En la sección 2.2 redujimos las posibles familias reales de polinomios ortogonales clásicos a tres casos distintos: las familias de Hermite, Laguerre y Jacobi. Tanto los polinomios de Laguerre $\{L_n^{(\alpha)}\}$, como los de Jacobi $\{P_n^{(\alpha,\beta)}\}$ dependen respectivamente de uno y dos parámetros, teniendo en todos los casos que cumplir que estos parámetros sean mayores que -1 .

Por otro lado, en la sección 2.3 hemos visto algunas de las propiedades más relevantes que caracterizan a los polinomios ortogonales clásicos. En esta sección, teniendo presente la fórmula de Rodrigues que se introdujo en la sección 2.3.3, presentaremos y describiremos algunas gráficas de las familias de Hermite, Laguerre y Jacobi para comprender el impacto sobre estas de la variación de los parámetros y del grado del polinomio. Todas las gráficas que se presentan a lo largo de esta sección (y algunas más), pueden consultarse en el archivo `software/graficas-clasicos.ipynb`³. En este archivo además también pueden encontrarse gráficas interactivas en las que pueden modificarse dinámicamente los parámetros de cada polinomio, permitiendo así observar los cambios en las gráficas de manera visual.

2.4.1. Polinomios de Hermite

En el caso de la familia de polinomios de Hermite, $\{H_n\}$, estos no dependen de ningún parámetro, por lo que tan sólo mostraremos la variación de sus gráficas respecto al grado del polinomio. Véanse las imágenes 2.2.

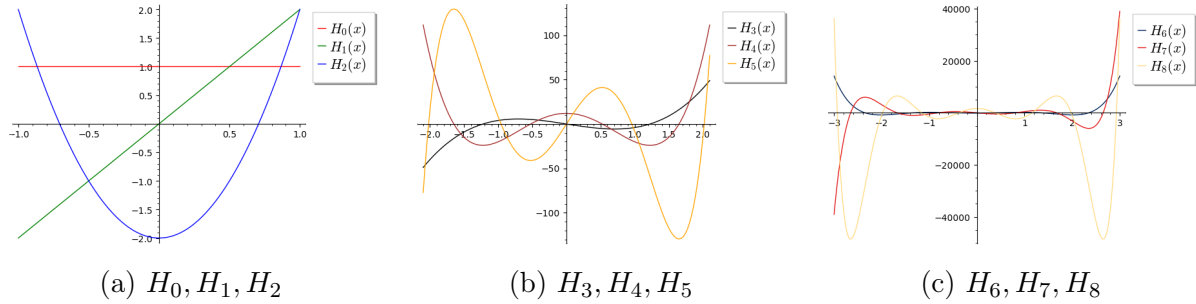


Imagen 2.2: Polinomios de Hermite

La primera evidencia es que los polinomios son pares cuando el grado n es par, e impares cuando el grado es impar. Por otra parte, nótese que el rango de representación del eje X escogido en cada una de las tres imágenes es cada uno mayor que el anterior. Esto se debe a que se ha buscado representar todos los ceros de cada polinomio en la gráfica, y recordemos que el intervalo $[x_{n1}, x_{nn}]$ es cada vez más grande conforme n aumenta (véase corolario 1.11). De hecho, si quisiéramos representar todos los ceros de todos los polinomios de Hermite sería imposible, pues el verdadero intervalo de ortogonalidad es $I = (-\infty, +\infty)$.

³Puede ser consultado en la URL <https://github.com/JAntonioVR/Polinomios-Ortogonales/blob/main/software/graficas-clasicos.ipynb>

2.4.2. Polinomios de Laguerre

Los polinomios de Laguerre, tal y como indica su notación, $\{L_n^{(\alpha)}\}$, admiten un parámetro real $\alpha > -1$. Además, su intervalo de ortogonalidad es $[0, +\infty)$, por lo que representaremos subconjuntos acotados de este intervalo. En las imágenes 2.3 podemos ver gráficas de los polinomios de Laguerre para distintos valores de α y un grado fijo. En particular, las imágenes 2.3(a) y 2.3(b) presentan gráficas para los polinomios de Laguerre de grado 2 y 3 respectivamente para valores de $\alpha = 0, 1, 5, 15, 20, 25$. El efecto que notamos es que a mayor valor de α la cola derecha es más suave y el crecimiento (resp. decrecimiento) hacia $+\infty$ (resp. $-\infty$) es más lento conforme crece α .

En la imagen 2.3(c) presentamos de nuevo polinomios de Laguerre de grado 3 pero con valores de α pequeños en valor absoluto, concretamente $\alpha = -0.9, -0.5, 0, 0.5, 1$. Vemos que la conclusión es la misma que en las gráficas (a) y (b) salvo cambio de escala.

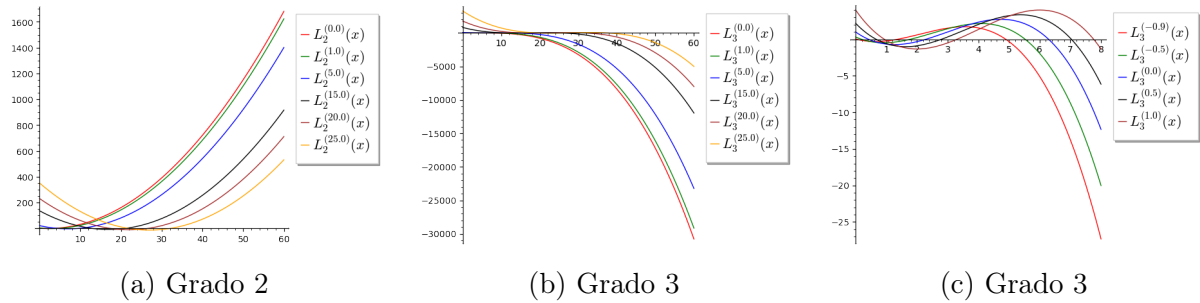


Imagen 2.3: Polinomios de Laguerre con distintos valores de α

Véamos ahora el efecto que tiene en las gráficas el aumento del grado n . Para ello, fijamos $\alpha = 0$ y observamos algunas imágenes en las que se ha modificado el grado. En la imagen 2.4(a) podemos observar las gráficas de grados 0, 1 y 2, mientras que en la imagen (b) se representan los polinomios de grados 3, 4 y 5 y en la (c) los grados 6, 7 y 8.

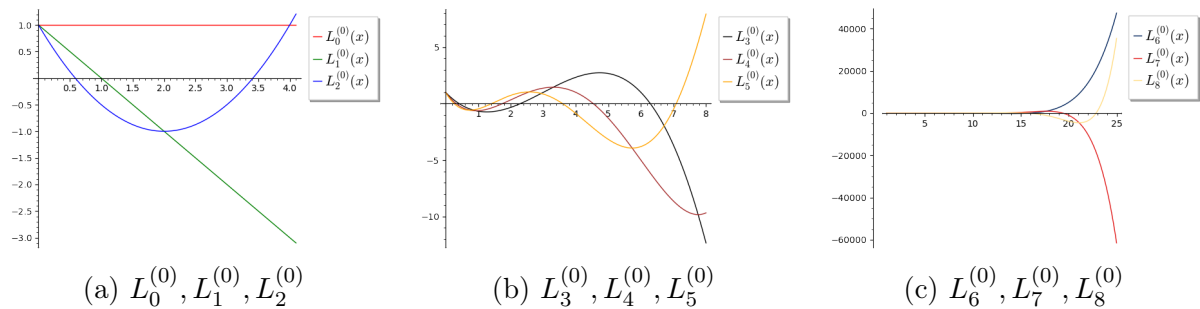


Imagen 2.4: Polinomios de Laguerre de distintos grados

En este caso no se presenta, al menos de manera evidente, ningún tipo de simetría o paridad como sí se manifestaba en el caso de los polinomios de Hermite. Por otro lado, nótese que la gráfica (b) está truncada, es decir, no aparecen todos los ceros de todos los polinomios. Por ejemplo, aunque sabemos que $L_5^{(0)}$ tiene 5 ceros simples, únicamente representamos 4 de ellos con el objetivo de observar mejor las ondulaciones y curvas de las gráficas. Sin embargo, en la gráfica (c) sí hemos representado todos los ceros de los

tres polinomios indicados, de forma que para que puedan visualizarse todos ellos en una misma gráfica se ha elegido el rango $[0, 25]$ y hemos renunciado a los detalles de las curvas en valores más pequeños de x .

Podemos recordar entonces la discusión sobre los ceros ya introducida en la sección anterior. Los intervalos $[x_{n1}, x_{nn}]$ son cada vez más amplios, de forma que el primer cero es cada vez más pequeño y cercano a cero, como puede verse en las imágenes 2.4 (a) y (b), mientras que el último cero crece sin límite. Esto se debe a que el verdadero intervalo de ortogonalidad en este caso es $[0, +\infty)$.

2.4.3. Polinomios de Jacobi

Por último, analizaremos los polinomios de Jacobi $\{P_n^{(\alpha, \beta)}\}$. Estos admiten hasta dos grados de libertad $\alpha, \beta > -1$ y manifiestan su ortogonalidad en el intervalo $[-1, 1]$, lo cual facilita considerablemente la representación gráfica de estos polinomios.

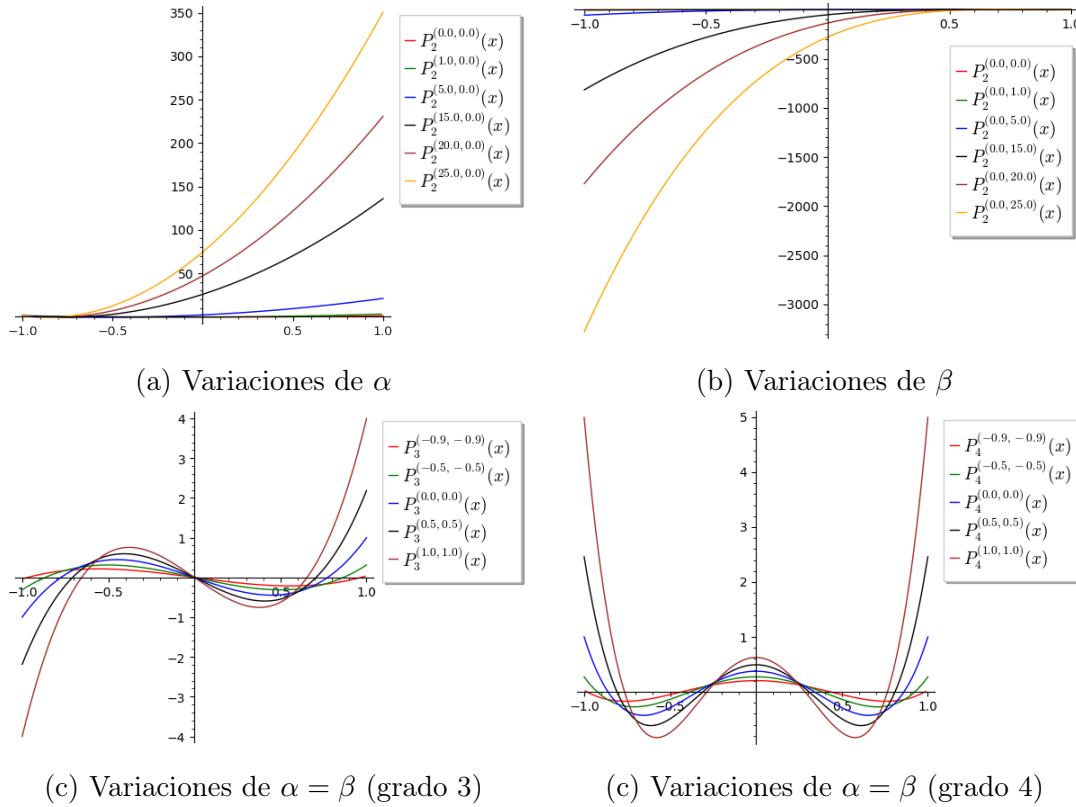


Imagen 2.5: Polinomios de Jacobi con distintos valores de α y β

Comenzaremos analizando el efecto del parámetro α . Para ello, hemos fijado $n = 2$ y $\beta = 0$, representando el polinomio de Jacobi $P_2^{(\alpha, 0)}$ para los valores $\alpha = 0, 1, 5, 15, 20, 25$, véase la imagen 2.5(a). Obsérvese que, de forma similar a los polinomios de Laguerre, este parámetro interviene en la velocidad a la que aumenta la cola derecha. Sin embargo, lo hace de forma inversa, ya que mientras que a mayores valores de α la inclinación era menor en los polinomios de Laguerre, en este caso conforme crece α aumenta también la misma.

Respecto a β , para analizar su impacto en las gráficas fijamos $n = 3$ y $\alpha = 0$,

presentando los polinomios $P_3^{(0,\beta)}$ para $\beta = 0, 1, 5, 15, 20, 25$ en la imagen 2.5(b). El comportamiento de β es análogo al de α pero en la cola izquierda en lugar de la derecha. Es decir, a mayor valor de β , mayor inclinación de la cola derecha, que en estas gráficas se traduce en un descenso a $-\infty$ más pronunciado.

En los casos de Hermite y Laguerre hemos discutido también la paridad y/o simetría de los gráficos. A juzgar por las imágenes 2.5(a) y 2.5(b), los polinomios de Jacobi no cumplen ningún tipo de simetría en general. Sin embargo, en la imagen 2.5(c) puede comprobarse qué ocurre al igualar $\alpha = \beta$. En efecto, en este caso las gráficas son simétricas respecto al eje Y cuando el grado es par ($P_{2n}^{(\alpha,\alpha)}$ es par). Respectivamente, $P_{2n+1}^{(\alpha,\alpha)}$ es impar.

Finalmente, comprobemos gracias a las imágenes 2.6 la variación de estas gráficas al aumentar el grado cuando fijamos $\alpha = \beta = 0$.

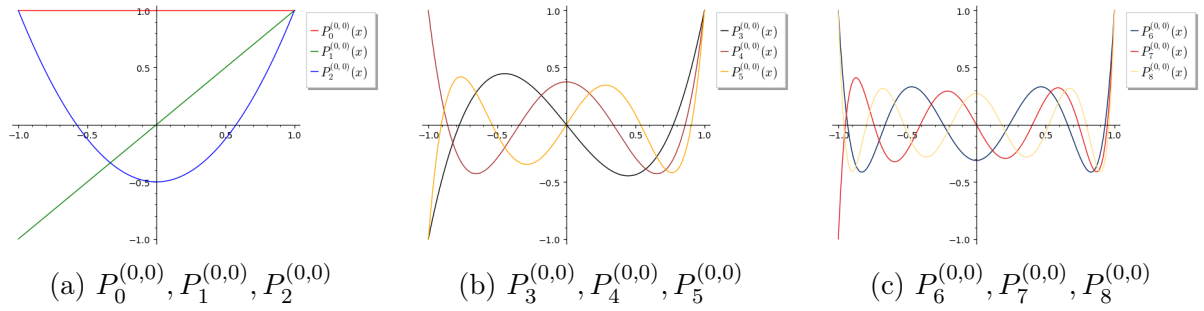


Imagen 2.6: Polinomios de Jacobi de distintos grados

En estas tres figuras se puede comprobar más fácilmente la paridad e imparidad de los polinomios de Jacobi en el caso $\alpha = \beta$ con respecto al grado n . Como además el intervalo de ortogonalidad es $I = [-1, 1]$, todos los ceros de estos polinomios están representados siempre en el mismo.

Hasta este momento, en los dos primeros capítulos, nos hemos centrado totalmente en una introducción teórico-práctica a los polinomios ortogonales y sus principales propiedades, profundizando en las llamadas familias clásicas. En este capítulo explicaremos una teoría probabilística que, aparentemente, nada tiene que ver con los polinomios ortogonales, aunque estudiaremos la estrecha relación que hay entre ambas teorías en el capítulo 4.

En concreto, analizaremos desde un punto de vista general los llamados *procesos de Markov*, *procesos markovianos* o simplemente *cadenas de Markov*. Estos modelos, basados en estados, tratan de explicar las probabilidades de los posibles cambios de estado de un sistema a partir de un principio denominado *propiedad de Markov*. Esta denominación tanto del modelo como de la propiedad se debe a ANDREY MARKOV, matemático que fue el primero en estudiar, en la década de 1900, los procesos estocásticos con dicha propiedad, que posteriormente se popularizarían.

Nos centraremos principalmente en procesos discretos, es decir, consideramos que el tiempo se divide en instantes $n = 0, 1, 2, \dots$, sin olvidar que existe una formulación análoga en la que el tiempo es continuo $t \in [0, \infty)$. Denominamos *estado* a cualquiera de las situaciones en las que se puede encontrar el sistema en un instante determinado. Por ejemplo, pensemos en un barrio en el que hay N fruterías y dividimos el tiempo en semanas. Asumimos que un individuo siempre va a comprar a una de estas tiendas aunque hay posibilidad de que cambie de una a otra distinta cada semana. Etiquetamos cada una de las fruterías con un número: frutería 0, frutería 1, \dots , frutería N .

3.1. Cadenas de Markov discretas. Matriz de transición

En general, consideraremos un espacio de estados $S \subseteq \mathbb{N}_0$ que con frecuencia denotaremos como $S = \{0, 1, \dots, N\}$ si consideramos un espacio finito de estados y $S = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ en caso de considerarlo infinito. A su vez, dividimos el tiempo en instantes $n = 0, 1, 2, \dots$. Fijado n , llamamos $X_n \in S$ a la variable aleatoria “estado del sistema en el instante n ”. La propiedad específica que consideró Markov consiste en

asumir que, suponiendo que el sistema se encuentra en un estado concreto en un instante concreto, la probabilidad de que ocurra una transición en el siguiente instante depende única y exclusivamente del estado actual.

Definición 3.1 (Cadena de Markov). El conjunto de variables aleatorias $\{X_n, n \geq 0\}$ sobre el espacio de estados S verifica la **propiedad markoviana** o **propiedad de Markov** si, para todo $n \in \mathbb{N}_0$ y $j, i, i_{n-1}, \dots, i_0 \in S$:

$$P[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0] = P[X_{n+1} = j | X_n = i]. \quad (3.1)$$

En caso de que $\{X_n, n \geq 0\}$ cumpla la propiedad (3.1) se denominará **cadena de Markov discreta**.

Una cadena de Markov $\{X_n, n \geq 0\}$ se llamará **homogénea** si, para todo $n \in \mathbb{N}_0$,

$$P[X_{n+1} = j | X_n = i] = P[X_1 = j | X_0 = i]. \quad (3.2)$$

Es decir, la probabilidad de que en un instante $n+1$ el sistema se encuentre en el estado j es independiente de lo que haya ocurrido en los instantes anteriores a n . Por ejemplo, aplicando esta propiedad al tiempo atmosférico podríamos afirmar que la probabilidad de que mañana llueva únicamente depende de si hoy ha llovido, sin tener en cuenta el clima de los días anteriores. En epidemiología, es como afirmar que el número de contagios de una enfermedad un día concreto puede predecirse a partir del número de contagios que haya habido el día anterior, independientemente del número de contagios de los días anteriores ni los posibles factores externos. Intuitivamente, no parece que esta suposición sea muy realista y no cabe esperar resultados precisos en predicciones futuras. Sin embargo, lo cierto es que es útil para hacernos una idea de lo que puede ocurrir a largo plazo con relativa exactitud y es muy utilizada a la hora de implementar medidas en ámbitos como la migración, la extinción de animales, uso de suelo, etc.

La probabilidad $P[X_{n+1} = j | X_n = i]$ se denomina **probabilidad de transición en un paso** de la cadena en el instante n . Si además la cadena es homogénea, este valor es el mismo para todo instante n , de ahí el término ‘homogénea’. En adelante, cuando utilicemos el término ‘cadena de Markov’ nos estaremos refiriendo a una cadena de Markov homogénea. Una vez anulada la dependencia de n , introducimos una nueva notación:

$$P_{ij} := P[X_{n+1} = j | X_n = i], \quad \forall i, j \in S.$$

Con esta notación, y teniendo en cuenta que en un espacio finito de $N+1$ estados existen $(N+1)^2$ posibles transiciones, agrupamos todas estas probabilidades en una matriz, que en el caso de que S sea infinito resultaría ser una matriz infinita.

Definición 3.2 (Matriz de transición). Llamamos **matriz de transición** de una cadena de Markov $\{X_n, n \geq 0\}$ a la matriz

$$P = \begin{pmatrix} P_{00} & P_{01} & \cdots & P_{0N} \\ P_{10} & P_{11} & \cdots & P_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N0} & P_{N1} & \cdots & P_{NN} \end{pmatrix} \quad \text{ó} \quad P = \begin{pmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ P_{20} & P_{21} & P_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (3.3)$$

(si S es finito) (si S es infinito)

Observación 3.1. Obsérvese que la i -ésima fila corresponde a la distribución de la probabilidad saliente del estado i , mientras que la j -ésima columna describe la probabilidad entrante al estado j .

Presentamos a continuación dos propiedades directas de las matrices de transición.

Proposición 3.1. Sea $\{X_n, n \geq 0\}$ una cadena de Markov sobre el espacio de estados $S \subseteq \mathbb{N}_0$ y cuya matriz de transición es P dada como en (3.3). Entonces

1. $0 \leq P_{ij} \leq 1 \quad \forall i, j \in S$.
2. Si $S = \{0, 1, \dots, N\}$, $\sum_{j=0}^N P_{ij} = 1 \quad \forall i \in S$ (todas las filas de P suman 1).
3. Si $S = \mathbb{N}_0$, $\sum_{j=0}^{\infty} P_{ij} = 1 \quad \forall i \in S$.

Demostración. La propiedad (1) es evidente teniendo en cuenta que P_{ij} es una probabilidad. Sobre la propiedad (2), tenemos que

$$\begin{aligned} \sum_{j=1}^N P_{ij} &= \sum_{j=1}^N P[X_{n+1} = j | X_n = i] \\ &= P[X_{n+1} \in S | X_n = i] \quad (\text{por el teorema de la probabilidad total}) \\ &= 1 \end{aligned}$$

Y la demostración de (3) es análoga a la de (2). □

A las matrices que cumplen estas dos propiedades se les denomina **matrices estocásticas** y pueden ser consideradas matrices de transición de alguna cadena de Markov.

Ejemplo 7 (Tiempo atmosférico). Extraído de [Kul12, Ch. 2, Example 2.3]. Supongamos que podemos clasificar el tiempo de una ciudad como soleado (0), nublado (1) o lluvioso (2) y que el tiempo del día siguiente únicamente depende del tiempo del día anterior con las siguientes probabilidades:

- Si hoy hace sol, mañana estará nublado con probabilidad 0.3 y lloverá con probabilidad 0.2.
- Si hoy está nublado, mañana hará sol con probabilidad 0.5 y lloverá con probabilidad 0.3.
- Si hoy está lloviendo, mañana será un día soleado con probabilidad 0.4 y nublado con probabilidad 0.5.

Modelaremos este sistema mediante una cadena de Markov. Para ello, consideraremos la variable $X_n \equiv$ “Tiempo atmosférico en el día n ”. Y consideramos el espacio de estados $S = \{0 \equiv \text{soleado}, 1 \equiv \text{nublado}, 2 \equiv \text{lluvioso}\}$. Atendiendo a las indicaciones anteriores y al ítem (2) de la proposición 3.1, la matriz de transición sería por tanto

$$P = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.5 & 0.2 & 0.3 \\ 0.4 & 0.5 & 0.1 \end{pmatrix}.$$

3.2. Distribuciones transitorias

Hasta el momento nos hemos centrado en la adaptación de distintas situaciones de acuerdo a modelos de estados regidos por cadenas de Markov, que en el caso homogéneo concentra toda su información en una matriz estocástica que denominamos matriz de transición. Sin embargo, ya adelantamos en la introducción que la principal función de este método consiste en tratar de predecir el comportamiento de la variable X_n en algún instante futuro. Pongamos el caso del ejemplo 7 sobre el tiempo atmosférico. Podría interesar predecir el tiempo que va a hacer dentro de 2 o 3 días para posponer o no un evento por lluvia. En esta sección estudiaremos cómo podemos utilizar cadenas de Markov con este objetivo. En adelante supondremos que el espacio de estados S es finito con $N+1$ elementos, es decir, $S = \{0, 1, \dots, N\}$.

Sea $\{X_n, n \geq 0\}$ una cadena de Markov discreta y homogénea sobre un espacio de estados $S = \{0, 1, \dots, N\}$, una matriz de transición P y supongamos conocida la distribución inicial, que denotaremos en forma de vector fila como $a^{(0)} = (a_0, \dots, a_N)$, donde $a_i = P[X_0 = i]$, $\forall i \in S$. Se denomina **distribución transitoria** a la distribución que sigue la variable X_n para cierto $n \geq 0$ fijo cuando es conocida la distribución de X_0 .

Abordemos este problema con lo que conocemos hasta el momento. Tenemos que

$$\begin{aligned} P[X_n = j] &= \sum_{i=0}^N P[X_n = j | X_0 = i] P[X_0 = i] \\ &= \sum_{i=0}^N a_i P[X_n = j | X_0 = i]. \end{aligned} \tag{3.4}$$

Por lo que nuestro problema se reduce a encontrar la probabilidad de que el proceso esté en el estado j tras n instantes suponiendo que inicialmente se encontraba en el estado i , $P[X_n = j | X_0 = i]$. Esta probabilidad extiende a la que en su momento llamamos probabilidad de transición en un paso y, naturalmente, se denomina **probabilidad de transición en n pasos**. Utilizamos la notación

$$a_i^{(n)} := P[X_n = i], \quad P_{ij}^{(n)} := P[X_n = j | X_0 = i],$$

$\forall i, j \in \{0, 1, \dots, N\}$. De esta forma, estamos denotando la función masa de probabilidad (a menudo abreviada como “fmp”) de la variable aleatoria X_n mediante el vector $a^{(n)} = (a_0^{(n)}, \dots, a_N^{(n)})$. Recordemos que la fmp de una variable aleatoria discreta determina unívocamente su distribución, por lo cual en adelante utilizaremos indistintamente el término ‘distribución’ y ‘función masa de probabilidad’, sin olvidar que en realidad son conceptos distintos.

Análogamente al caso en el que considerábamos un único paso, agrupamos todos los posibles valores de $P_{ij}^{(n)}$ en una matriz.

Definición 3.3 (Matriz de transición en n pasos). Llamamos **matriz de transición en**

n pasos de una cadena de Markov $\{X_n, n \geq 0\}$ a la matriz

$$P^{(n)} = \begin{pmatrix} P_{00}^{(n)} & P_{01}^{(n)} & \cdots & P_{0N}^{(n)} \\ P_{10}^{(n)} & P_{11}^{(n)} & \cdots & P_{1N}^{(n)} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N0}^{(n)} & P_{N1}^{(n)} & \cdots & P_{NN}^{(n)} \end{pmatrix}. \quad (3.5)$$

Observación 3.2. Si tomamos $n = 0$, tenemos que $P_{ij}^{(0)} = \delta_{ij}$, $i, j \in S$. Esto es, $P^{(0)} = I_{N+1}$. Por otro lado, si $n = 1$, tenemos que $P^{(1)} = P$, la matriz de transición (3.3).

A partir de la definición de matriz de transición en n pasos podemos reescribir (3.4) de forma matricial como

$$a^{(n)} = a^{(0)} \cdot P^{(n)}.$$

Por lo que resta obtener un método para calcular la matriz de transición en n pasos, $P^{(n)}$. Este método nos lo da el siguiente teorema:

Teorema 3.1 (Matriz de transición en n pasos). *Sea $\{X_n, n \geq 0\}$ una cadena de Markov homogénea con matriz de transición P , entonces*

$$P^{(n)} = P^n. \quad (3.6)$$

Demostración. Consúltese [Kul12, Ch. II, theorem 2.2]. □

De este resultado obtenemos dos sencillas consecuencias:

Corolario 3.2. *La variable aleatoria X_n se distribuye según la función masa de probabilidad descrita por el vector*

$$a^{(n)} = a^{(0)} \cdot P^n.$$

Y aplicando la homogeneidad que asumimos en toda cadena:

Corolario 3.3.

$$P[X_{n+m} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0] = P[X_{n+m} = j | X_n = i] = P_{ij}^{(m)},$$

$$j, i, i_{n-1}, \dots, i_0 \in S.$$

Siguientemente, presentamos una ecuación importante en este ámbito.

Teorema 3.4 (Ecuación de Chapman-Kolmogorov). *Sea $\{X_n, n \geq 0\}$ una cadena de Markov discreta y homogénea, entonces las probabilidades de transición en n pasos satisfacen la siguiente igualdad, denominada ecuación de Chapman-Kolmogorov.*

$$P_{ij}^{(n+m)} = \sum_{k=0}^N P_{ik}^{(n)} P_{kj}^{(m)}. \quad (3.7)$$

Demostración.

$$\begin{aligned}
P_{ij}^{(n+m)} &= P[X_{n+m} = j | X_0 = i] \\
&= \sum_{k=0}^N P[X_{n+m} = j | X_n = k, X_0 = i] P[X_n = k | X_0 = i] \text{ (th. de la prob. total)} \\
&= \sum_{k=0}^N P[X_{n+m} = j | X_n = k] P[X_n = k | X_0 = i] \text{ (corolario 3.3)} \\
&= \sum_{k=0}^N P[X_m = j | X_0 = k] P[X_n = k | X_0 = i] \text{ (por la homogeneidad)} \\
&= \sum_{k=0}^N P_{ik}^{(n)} P_{kj}^{(m)}
\end{aligned}$$

□

Si escribimos la ecuación (3.7) en forma matricial, tenemos que

$$P^{(n+m)} = P^{(n)} \cdot P^{(m)},$$

lo cual es evidente teniendo en cuenta el teorema 3.1. Intercambiando los roles de n y m llegamos a la conclusión de que las matrices $P^{(n)}$ y $P^{(m)}$ conmutan para todo $n, m \geq 0$. El teorema 3.1 es la clave del método para calcular las distribuciones transitorias, pues reduce el problema a potencias y multiplicaciones de matrices.

Ejemplo 8. Tratemos ahora un ejemplo distinto. En este caso nos centraremos en las variaciones interprovinciales. La tabla 3.1 muestra en cada celda el número de individuos que cambiaron su vivienda habitual desde la provincia indicada en la fila hacia la provincia indicada en la columna correspondiente (datos extraídos del INE: Instituto Nacional de Estadística).

Provincia	Alicante	Castellón	Valencia
Alicante	41.685	484	3.957
Castellón	369	11.059	2.232
Valencia	3.516	2.903	71.741

Tabla 3.1: Variaciones interprovinciales en la comunidad valenciana (2021)

Por su parte, la tabla 3.2 muestra la población de cada una de las provincias en el año 2021.

Provincia	Población
Alicante	1881762
Castellón	587064
Valencia	2589312

Tabla 3.2: Población de las provincias de la comunidad valenciana (2021)

Con estos datos y asumiendo que la población total permanece constante, ¿podemos comprobar la población de cada una de las provincias de la Comunidad Valenciana en el año 2022? ¿Y predecir las que habrá en el año 2023?

Antes de proceder al propio modelado, queremos mencionar que estos datos han sido manipulados y presentados utilizando el software estadístico R¹. El archivo de ‘R Markdown’ `software/ejemplo-migraciones.Rmd` contiene el código fuente en R y algunas descripciones, mientras que el fichero `software/ejemplo-migraciones.html`, aunque tiene el mismo contenido, es un archivo compilado que es posible abrir con cualquier navegador web y en el que no es necesario ejecutar nada para visualizar las salidas de los distintos comandos².

Procedemos a modelar el problema como una cadena de Markov $\{X_n, n \geq 0\}$, donde $X_n \equiv$ Provincia en la que reside un individuo al azar de la C. Valenciana en $2021 + n$. Esta cadena se construye sobre los estados

$$S = \{0 \equiv \text{‘Alicante’}, 1 \equiv \text{‘Castellón’}, 2 \equiv \text{‘Valencia’}\}.$$

Podemos tomar una matriz asociada a la tabla 3.1 y a partir de ella construir la matriz de transición de la cadena, teniendo en cuenta que de esta matriz la celda (i, j) representa el número de migraciones de la provincia i a la provincia j . Sin embargo, es claro que la tabla 3.1 no verifica las condiciones de la proposición 3.1, por lo que no podemos formar directamente una matriz estocástica. De igual manera, el vector cuyas componentes son las de la tabla 3.2 no puede denotar una función masa de probabilidad pues sus componentes no suman 1. Por suerte, estos problemas pueden solucionarse rápidamente sin más que dividir cada fila de la tabla 3.1 entre el total de cada fila y cada componente de la tabla 3.2 entre la suma de las tres celdas. De esta forma, tenemos la siguiente matriz de transición P y la distribución inicial $a^{(0)}$.

$$P = \begin{pmatrix} 0.90372024 & 0.01049300 & 0.08578676 \\ 0.02701318 & 0.80959004 & 0.16339678 \\ 0.04498465 & 0.03714176 & 0.91787359 \end{pmatrix},$$

$$a^{(0)} = (0.3720266 \quad 0.1160633 \quad 0.5119101).$$

Por tanto, para conocer la distribución poblacional del año 2022 y la de 2023 simplemente tenemos que aplicar el teorema 3.1 y el corolario 3.2 para $n = 1$ y $n = 2$. Así,

$$X_1 \sim a^{(1)} = a^{(0)} \cdot P = (0.3623713 \quad 0.1168806 \quad 0.5207481),$$

$$X_2 \sim a^{(2)} = a^{(0)} \cdot P^2 = (0.3540653 \quad 0.1177692 \quad 0.5281655).$$

Por último, si lo que queremos es predecir la población (y no realmente su distribución), tendríamos que multiplicar estas distribuciones por la población total en 2021, obteniendo las poblaciones que recogemos en la tabla 3.3.

¹Consúltase su página web <https://www.r-project.org/>.

²Ambos ficheros están disponibles en las URLs <https://github.com/JAntonioVR/Polinomios-Ortogonales/blob/main/software/ejemplo-migraciones.Rmd>, <https://github.com/JAntonioVR/Polinomios-Ortogonales/blob/main/software/ejemplo-migraciones.html>

Provincia	2022	2023
Alicante	1 832 924	1 790 911
Castellón	591 198.1	595 692.9
Valencia	2 634 016	2 671 534

Tabla 3.3: Predicciones de la población de las provincias de la C. Valenciana en 2022 y 2023

3.3. Comportamiento límite

De nuevo, sea $\{X_n, n \geq 0\}$ una cadena de Markov homogénea sobre un conjunto finito de estados $S = \{0, 1, \dots, N\}$. En las secciones anteriores hemos aprendido a calcular la distribución de la variable X_n para cualquier $n \in \mathbb{N}_0$, pero es el momento de preguntar qué ocurre con dicha distribución cuando $n \rightarrow \infty$. Surgen así varias preguntas: ¿La distribución de X_n converge a alguna distribución límite? ¿En caso de que converja, la distribución límite es única? ¿Cómo podríamos calcularla? En esta sección trataremos, de manera resumida, de responder a dichas preguntas.

3.3.1. Clasificación de los estados de una cadena de Markov

Antes de enunciar los principales resultados sobre existencia y unicidad de límite introducimos alguna terminología sobre las cadenas de Markov y sus estados.

Definición 3.4. Sean $i, j \in S$ dos estados. Decimos que el estado j es **accesible** desde el estado i si existe algún $k \geq 0$ tal que $P_{ij}^{(k)} > 0$. Decimos que los estados i y j **se comunican** o que son **comunicantes**, y lo denotamos como $i \leftrightarrow j$, si son accesibles entre ellos.

La siguiente proposición nos permitirá trabajar con cierta comodidad con la relación de comunicación.

Proposición 3.2. La ‘comunicación’ (\leftrightarrow) es una relación de equivalencia sobre S .

Demostración. Consúltase [Ros83, Proposition 4.2.1]. □

Definición 3.5 (Cadena de Markov irreducible). Una cadena de Markov $\{X_n, n \geq 0\}$ sobre un conjunto de estados S es **irreducible** si todos sus estados se comunican entre sí, es decir, si el cociente S/\leftrightarrow consta de una sola clase de equivalencia. Una cadena de Markov que no es irreducible se dice **reducible**.

Nótese que la condición de cadena irreducible equivale a que sea posible llegar a cualquier estado j desde cualquier estado i . Podemos comprobarlo con la matriz de transición sin más que mirar las entradas correspondientes, pero existe un mecanismo más visual e intuitivo que la matriz de transición: el **diagrama de transición**.

Si una cadena de Markov $\{X_n, n \geq 0\}$ sobre un conjunto de estados $S = \{1, \dots, N\}$ tiene una matriz de transición P , podemos construir su diagrama de transición de la siguiente forma:

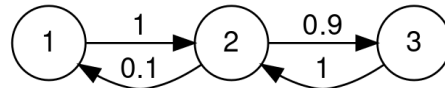
1. Dibujamos un nodo por cada posible estado, cada uno con su correspondiente etiqueta.
2. Para cada $i, j \in S$, trazamos una flecha desde el nodo i, j únicamente si la entrada (i, j) de la matriz es estrictamente positiva. Esto es, si $P_{ij} > 0$. En caso de dibujar la flecha, es recomendable etiquetarla con el correspondiente valor de P_{ij} .

Entonces, podemos trasladar las anteriores definiciones con respecto al diagrama de Markov:

- Un estado j es accesible desde el estado i si, y solo si existe un camino en el diagrama de transición desde el nodo i hacia el nodo j .
- Dos estados i, j son comunicantes si, y solo si existe un camino del nodo j al i y viceversa.
- Una cadena es irreducible si, y solo si cada pareja de nodos es mutuamente accesible.

Ejemplo 9. 1. Supongamos una cadena de Markov cuya matriz y diagrama de transición son los siguientes:

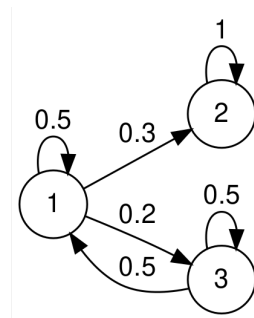
$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0.1 & 0 & 0.9 \\ 0 & 1 & 0 \end{pmatrix}$$



En este caso, la cadena sería irreducible.

2. Supongamos ahora que la matriz y el diagrama son:

$$P = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0 & 1 & 0 \\ 0.5 & 0 & 0.5 \end{pmatrix}$$



Fijémonos que, aunque el estado 2 es accesible desde todos los estados, ni el estado 1 ni el 3 son accesibles desde el 2, lo cual impide que la cadena sea irreducible. Cuando un estado cumple esta propiedad se denomina estado *absorbente*, término que definiremos formalmente más adelante.

Hemos clasificado los estados según su accesibilidad, pero fijémonos ahora en la posibilidad de visitar o no un estado. En el ejemplo 9.1 es posible visitar cualquiera de los tres estados cada dos transiciones, mientras que en el ejemplo 9.2 es posible mantenerse en cualquiera de los tres estados iteración tras iteración. Sobre esta noción insiste la siguiente definición.

Definición 3.6 (Estado periódico, aperiódico y absorbente). Un estado i es **periódico** de periodo d_i si la cadena puede regresar a i en pasos múltiplos de d_i . Si $d_i = 1$, decimos que el estado i es **aperiódico**. Si ningún estado de la cadena es accesible desde el estado i (i.e. $P_{ii} = 1$), entonces el estado i se denomina **absorbente**.

Observación 3.3. Si el estado j es comunicante con un estado periódico, entonces j es periódico.

Podemos extender esta definición a la cadena de Markov completa:

Definición 3.7 (Cadena de Markov periódica y aperiódica). Sea $\{X_n, n \geq 0\}$ una cadena de Markov homogénea, irreducible sobre un espacio de estados $S = \{0, \dots, N\}$. Sea d el mayor número entero tal que,

$$P_{ii}^{(n)} > 0 \Rightarrow d|n, \quad \forall i \in S. \quad (3.8)$$

Si $d > 1$, la cadena de Markov es **periódica**, si $d = 1$ es **aperiódica**.

Nótese que una cadena periódica puede volver a su estado inicial en $d, 2d, 3d, \dots$ transiciones. Lo interesante de esta definición y de que únicamente se aplique a cadenas irreducibles es que basta con hallar el mayor periodo d_i (según la definición 3.6) común a todos los estados, de forma que la irreducibilidad de la cadena garantiza que ese valor d es válido para todos los estados. Por otro lado, si $P_{ii} > 0$ para algún estado $i \in S$ en una cadena irreducible, necesariamente $d = 1$ y la cadena es aperiódica.

Observación 3.4. La periodicidad también es fácil de consultar en los diagramas. Basta con definir el ciclo más corto que empiece y termine en cada uno de los nodos. Entonces d es el mínimo común múltiplo de las longitudes de estos ciclos.

Así, la cadena del ejemplo 9.1 es periódica con $d = 2$. Sin embargo, no podemos plantearnos la periodicidad de la cadena del ejemplo 9.2, ya que ésta no es irreducible.

3.3.2. Teoremas límite

Una vez introducida la terminología principal de estados y cadenas, podemos estudiar la existencia y unicidad de una distribución límite. En primer lugar, vamos a asumir que existe al menos una distribución de probabilidad a la cual converge la distribución de X_n .

Definición 3.8. Llamamos **distribución límite** a la función masa de probabilidad a la cual converge la fmp de X_n cuando n tiende a infinito. Denotamos esta fmp como

$$\pi = (\pi_0, \dots, \pi_N), \quad (3.9)$$

donde $\pi_j = \lim_{n \rightarrow \infty} P[X_n = j]$, $j \in S$.

Si asumimos la existencia de una distribución límite, esta debe cumplir ciertas propiedades.

Proposición 3.3. Si una cadena de Markov $\{X_n, n \geq 0\}$ admite una distribución límite π , entonces ésta verifica las siguientes propiedades:

$$\pi_j = \sum_{i=0}^N \pi_i P_{ij}, \quad \forall j \in S \quad (3.10)$$

$$\sum_{j=0}^N \pi_j = 1 \quad (3.11)$$

Demostración. Aplicando el teorema de la probabilidad total y el teorema de la probabilidad condicionada, tenemos que

$$P[X_{n+1} = j] = \sum_{i=0}^N P[X_n = i] P_{ij}, \quad \forall j \in S. \quad (3.12)$$

Si tomamos límite en ambos miembros de la igualdad cuando $n \rightarrow \infty$ y asumimos la existencia de esta distribución límite obtenemos la igualdad (1). La igualdad (2) se asume de la representación de π como una función masa de probabilidad. \square

La ecuación (3.10) también se escribe en forma matricial como

$$\pi = \pi \cdot P. \quad (3.13)$$

Esta ecuación matricial puede resolverse, y si normalizamos el vector obtenido de forma que la suma de sus componentes resulte 1, obtenemos un vector que denominaremos π^* . Supongamos ahora que $a^{(0)} = \pi^*$. Entonces, aplicando que es solución de (3.13), $a^{(1)} = \pi^* \cdot P = \pi^*$, $a^{(2)} = \pi^* \cdot P^2 = (\pi^* \cdot P)P = \pi^* \cdot P = \pi^*$, ..., $a^{(n)} = \pi^*$. Es decir, la distribución de X_n , representada por el vector $a^{(n)}$, es la misma siempre independientemente de n e igual a π^* . Es lo que en adelante denominaremos distribución estacionaria.

Definición 3.9 (Distribución estacionaria). Una función masa de probabilidad dada por un vector fila $\pi^* = (\pi_0^*, \dots, \pi_N^*)$, se denomina **distribución estacionaria** de una cadena de Markov si

$$P[X_0 = i] = \pi_i^* \quad \forall i \in S \Rightarrow P[X_n = i] = \pi_i^* \quad \forall i \in S, n \geq 0 \quad (3.14)$$

Como vemos, las distribuciones límite y las distribuciones estacionarias guardan ambas una estrecha relación con las ecuaciones de la proposición 3.3, y vemos que también la tendrán la existencia y unicidad de una y otra.

Proposición 3.4. Una función masa de probabilidad dada por un vector fila $\pi^* = (\pi_0^*, \dots, \pi_N^*)$ es una distribución estacionaria si, y solo si verifica las ecuaciones (3.10) y (3.11).

Demostración. Lo probaremos por doble implicación.

\Rightarrow Supongamos que π^* cumple la propiedad de la definición 3.9. Si tomamos $n = 0$ en (3.12) obtenemos

$$P[X_1 = j] = \sum_{i=1}^N P[X_0 = i] P_{ij}, \quad \forall j \in S.$$

Pero como $P[X_0 = j] = \pi_j^*$ y $P[X_1 = j] = \pi_j^*$, sustituyendo obtenemos la igualdad (3.10). La ecuación (3.11) se deduce de que partimos de que π^* es una función masa de probabilidad representada en un vector.

\Leftarrow Ahora asumimos que π^* verifica las ecuaciones (3.10) y (3.11). Consideramos entonces que $P[X_0 = j] = \pi_j^*$, $j \in S$. De nuevo, tomando $n = 0$ en (3.12):

$$P[X_1 = j] = \sum_{i=1}^N P[X_0 = i] P_{ij} = \sum_{i=1}^N \pi_i^* P_{ij} \underbrace{=}_{(3.10)} \pi_j^*.$$

Por lo que la fmp de X_1 es π^* . Si usamos repetidamente (3.12) con distintos valores de n , deducimos que la fmp de X_n es π^* para cualquier $n \geq 0$. Esto es, π^* es una distribución estacionaria.

□

El hecho de que tanto distribuciones límite como distribuciones estacionarias cumplan las mismas ecuaciones tiene una consecuencia sencilla de comprobar.

Corolario 3.5. *Si una cadena admite una distribución límite, entonces ésta es también una distribución estacionaria.*

El concepto de cadena irreducible y el de distribución estacionaria confluyen en el siguiente resultado, cuya demostración omitiremos.

Teorema 3.6. *Toda cadena de Markov finita e irreducible admite una única distribución estacionaria, i.e., la ecuación (3.13) tiene una única solución normalizada.*

Ejemplo 10. La distribución $\pi_1 = (0.05, 0.5, 0.45)$ es estacionaria para la cadena del ejemplo 9.1.

Finalmente, presentaremos el último resultado de este capítulo, que es el que determina el comportamiento límite de las cadenas de Markov.

Teorema 3.7. *Toda cadena de Markov finita, irreducible y aperiódica admite una única distribución límite que es, además, estacionaria.*

Con este último teorema, sumado a la última afirmación del enunciado del teorema 3.6, podemos conocer rápidamente el comportamiento límite de las cadenas de Markov que sean irreducibles y aperiódicas sin más que encontrar la única solución de las ecuaciones (3.10) y (3.11) (o la solución normalizada de la ecuación matricial (3.13)).

El caso de cadenas de Markov reducibles y/o periódicas es más complejo. Por ejemplo, la distribución de X_n puede converger a un ciclo de longitud d si la cadena $\{X_n, n \geq 0\}$

es irreducible pero periódica con periodo d . O si es reducible, la distribución límite puede ser dependiente del estado inicial. Para un estudio más detallado de estos casos, así como la demostración de estos últimos teoremas, remitimos al lector a [Ros83, Ch. II, Sections 4.2 and 4.3].

Ejemplo 11. Retomamos el ejemplo 8 con la intención de predecir la población de cada provincia de la comunidad valenciana a largo plazo. Si nos fijamos en la matriz de transición podemos ver rápidamente que la cadena de Markov $\{X_n, n \geq 0\}$ es irreducible y aperiódica. Si resolvemos la ecuación $\pi = \pi \cdot P$, que es equivalente a resolver el sistema de ecuaciones

$$\pi_{1 \times 3}(P - I_3)_{3 \times 3} = 0_{1 \times 3}$$

obtenemos que su única solución normalizada es

$$\pi = \begin{pmatrix} 0.3021799 & 0.1278348 & 0.5699852 \end{pmatrix}$$

De esta forma, el modelo predice que la población a largo plazo de las provincias de la comunidad valenciana es la recogida en la tabla 3.4.

Provincia	Población
Alicante	1 528 468
Castellón	646 606.3
Valencia	2 883 064

Tabla 3.4: Predicción de la población de las provincias de la C. Valenciana en el futuro

Para finalizar este capítulo, aclaramos que todos estos teoremas y definiciones tienen un análogo en el que se trabaja con columnas en lugar de con filas. Es decir, existe una manera distinta aunque equivalente de formular la teoría de las cadenas de Markov en la que las columnas de las matrices de transición suman 1 (en lugar de las filas), las fmp se denotan con vectores columnas, las distribuciones transitorias se calculan multiplicando por la derecha un vector columna por la matriz de transición, etc. Es frecuente encontrar este enfoque en la literatura, aunque en nuestro caso se ha elegido éste por la intuición fácilmente recordable de que una probabilidad p_{ij} sea la ‘probabilidad de ir desde el estado i al estado j ’ por encima de que fuera la ‘probabilidad de llegar al estado i desde el estado j ’.

CAPÍTULO 4

POLINOMIOS ORTOGONALES Y PROCESOS DE NACIMIENTO Y MUERTE

En el capítulo anterior hemos conocido a los procesos de Markov discretos, más frecuentemente denominados *cadena de Markov*. Por otra parte, en los dos primeros capítulos hemos visto la teoría fundamental de polinomios ortogonales, profundizando en los conocidos polinomios clásicos. Aparentemente, son dos ámbitos completamente diferentes: los polinomios ortogonales están muy relacionados con teoría de la medida, álgebra lineal y análisis funcional, mientras que las cadenas de Markov son modelos meramente probabilísticos. A pesar de ello, ambas disciplinas confluyen en un caso particular de cadena de Markov: los procesos de nacimiento y muerte (*birth-and-death processes*).

Los procesos de nacimiento y muerte discretos, también llamados cadenas de nacimiento y muerte, caminatas aleatorias, o en inglés más conocidos como *random walks*, son cadenas de Markov en la que desde un determinado estado i tan solo es posible acceder al siguiente ($i + 1$), al anterior ($i - 1$), o permanecer en el mismo estado (i). Fueron los matemáticos Samuel Karlin (1924–2007) y James McGregor (1921–1988) quienes profundizaron en la relación de estas particulares cadenas con los polinomios ortogonales, destacando la que llamaremos fórmula de representación de Karlin y McGregor.

Para la redacción de este capítulo se han usado como bibliografía fundamental las referencias [Dom22, Ch. II, III], [Sch00, Ch. II], [KM59], [KM58], [KM57], entre otras.

4.1. Procesos de nacimiento y muerte

Como mencionamos anteriormente, un proceso de nacimiento y muerte discreto, cadena de nacimiento y muerte o caminata aleatoria es un caso particular de proceso de Markov en el que la única transición posible desde un estado $i \in S$ es hacia el estado siguiente, $i + 1$, hacia el anterior, $i - 1$, o hacia el propio estado i . Intuitivamente, es como un juego de plataformas infinito. Entonces, tomamos $S = \mathbb{N}_0 = \{0, 1, 2, \dots\}$ ¹ y, por esta suposición,

¹Existe una extensión en la que $S = \{-1, 0, 1, 2, \dots\}$ y el estado -1 es un estado absorbente. Nosotros supondremos $S = \mathbb{N}_0$, aunque si se desea se puede consultar [Sch00, Ch. II, Section 2.3] para más información.

la matriz del proceso sería la matriz tridiagonal infinita

$$P = \begin{pmatrix} r_0 & p_0 & 0 & 0 & \cdots \\ q_1 & r_1 & p_1 & 0 & \cdots \\ 0 & q_2 & r_2 & p_2 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}, \quad (4.1)$$

donde

1. $q_i = p_{i-1,i} = P[X_{n+1} = i-1 | X_n = i]$, $i = 1, 2, \dots$
2. $r_i = p_{i,i} = P[X_{n+1} = i | X_n = i]$, $i = 0, 1, 2, \dots$
3. $p_i = p_{i+1,i} = P[X_{n+1} = i+1 | X_n = i]$, $i = 0, 1, 2, \dots$
4. Asumimos que $q_i > 0$ ($i \geq 1$), $r_i \geq 0$ ($i \geq 0$), $p_i > 0$ ($i \geq 0$), $r_0 + p_0 = 1$, $q_i + r_i + p_i = 1$ ($i \geq 1$).

Es decir, siempre es posible avanzar o retroceder, permitiendo que no siempre sea posible mantenerse en el mismo estado. En la imagen 4.1 se representa el diagrama de estados asociado a esta particular cadena.

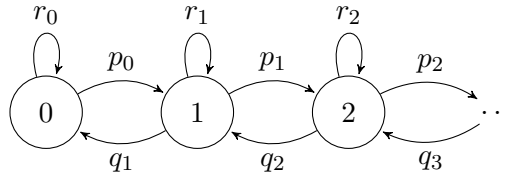


Imagen 4.1: Diagrama de transición de un proceso de nacimiento y muerte

El comportamiento a largo plazo de los procesos de nacimiento y muerte puede explicarse a partir de unos coeficientes que definimos a continuación:

Definición 4.1 (Coeficientes potenciales). Dado un proceso de nacimiento y muerte $\{X_n : n \geq 0\}$ con matriz de transición P definida como en (4.1), se definen sus **coeficientes potenciales** como

$$\pi_0 = 1 \quad \pi_n = \frac{p_0 p_1 \cdots p_{n-1}}{q_1 q_2 \cdots q_n} \quad (n \geq 1) \quad (4.2)$$

Nótese que, para $n \geq 1$ los coeficientes potenciales cumplen la relación de recurrencia $\pi_n = \pi_{n-1} \frac{p_{n-1}}{q_n}$, de la cual podemos deducir las siguientes ecuaciones de simetría:

$$\pi_n p_n = \pi_{n+1} q_{n+1} \quad (n \geq 0). \quad (4.3)$$

Por otro lado, obsérvese que, con las condiciones impuestas sobre P , el vector fila infinito $\pi = (\pi_n)_{n \in \mathbb{N}_0}$ verifica $\pi P = \pi$, la ecuación (3.13) extendida a vectores y matrices infinitos. Por tanto, π es un vector estacionario de la cadena, que de hecho será una distribución estacionaria si $\sum_{n=0}^{\infty} \pi_n < \infty$.

Por la estructura que posee la matriz P , podemos comprobar que siempre es irreducible y es aperiódica si existe algún $r_n > 0$. Bajo estas hipótesis, podemos pensar que es

posible aplicar el teorema 3.7 para asegurar la existencia de una distribución límite, pero recordemos que en este caso estamos tratando con cadenas de Markov infinitas. Afortunadamente, en [Sen81, Theorem 5.5] y en [Dom22, Theorem 2.3] podemos encontrar generalizaciones del teorema 3.7 que nos ayudan a deducir que el vector infinito π representa la distribución límite del proceso de nacimiento y muerte.

4.2. La fórmula de representación de Karlin-McGregor

En la última sección hemos presentado las cadenas de nacimiento y muerte desde un punto de vista probabilístico, extendiendo lo aprendido en el capítulo 3. En esta y en las próximas secciones estudiaremos qué relación existe entre estas particulares cadenas y los polinomios ortogonales. Para ello nos ayudaremos de varias técnicas analíticas, entre las que destacamos el conocido teorema espectral (teorema B.1), presentado con mayor nivel de detalle en el apéndice B.

Recordemos que las cadenas de nacimiento y muerte son cadenas de Markov infinitas en las que desde el estado i -ésimo tan solo se puede retroceder al estado anterior con probabilidad q_i , mantenerse en el mismo con probabilidad r_i o avanzar al siguiente con probabilidad p_i , obteniendo una matriz del proceso como la presentada en (4.1).

A partir de los coeficientes $\{q_n\}_{n \geq 1}$, $\{r_n\}_{n \geq 0}$ y $\{p_n\}_{n \geq 0}$ definimos recurrentemente la sucesión de polinomios $\{Q_n\}$ como

$$\begin{cases} Q_{-1}(x) = 0, & Q_0(x) = 1, \\ xQ_0(x) = p_0Q_1(x) + r_0Q_0(x), \\ xQ_n(x) = p_nQ_{n+1}(x) + r_nQ_n(x) + q_nQ_{n-1}(x), & (n \geq 1). \end{cases} \quad (4.4)$$

Nótese la similitud entre estas ecuaciones y la relación de recurrencia a tres términos (1.15) de los polinomios ortogonales.

Consideramos el espacio de Hilbert $\ell_\pi^2(\mathbb{N}_0)$, definido como el espacio de sucesiones complejas $f = \{f_n\}_{n \geq 0}$ tales que

$$\|f\|_\pi^2 = \sum_{n=0}^{\infty} |f_n|^2 \pi_n < \infty,$$

donde los coeficientes π_n son los dados en (4.2). Para cuando sea necesario, asumimos que $f_n = 0$ siempre que $n < 0$.

Teniendo presente la matriz $P = (P_{ij})_{i,j \geq 0}$ dada en (4.1), en el espacio de Hilbert $\ell_\pi^2(\mathbb{N}_0)$, asumiendo un abuso de notación consideramos la transformación

$$f \longmapsto Pf, \quad (4.5)$$

donde

$$(Pf)_n = \sum_{m=0}^{\infty} P_{nm}f_m = p_nf_{n+1} + r_nf_n + q_nf_{n-1}.$$

Obsérvese que en realidad Pf es el resultado de multiplicar la matriz P por f visto como un vector columna.

El objetivo es aplicar el teorema espectral a un operador que, aunque no será el propio P , estará muy relacionado con él. Primero nos aseguraremos de que P verifica las hipótesis del teorema espectral.

Lema 4.1. El operador P definido en (4.5) es un operador lineal, acotado y autoadjunto sobre el espacio de Hilbert $\ell_\pi^2(\mathbb{N}_0)$ cuya norma verifica $\|P\| \leq 1$.

Demostración. Sean $f, g \in \ell_\pi^2(\mathbb{N}_0)$.

- $Pf \in \ell_\pi^2(\mathbb{N}_0)$: Para asegurar que Pf pertenece al espacio $\ell_\pi^2(\mathbb{N}_0)$, podemos comprobar que las sucesiones $\{p_n f_{n+1}\}_{n \geq 0}$, $\{r_n f_n\}_{n \geq 0}$ y $\{q_n f_{n-1}\}_{n \geq 0}$ son elementos de $\ell_\pi^2(\mathbb{N}_0)$, de forma que Pf también lo será por ser suma de elementos de $\ell_\pi^2(\mathbb{N}_0)$. Así, tenemos que:

$$\begin{aligned} \sum_{n=0}^{\infty} |p_n f_{n+1}|^2 \pi_n &= \sum_{n=0}^{\infty} p_n^2 |f_{n+1}|^2 \pi_n \\ &= \sum_{n=0}^{\infty} p_n |f_{n+1}|^2 q_{n+1} \pi_{n+1} \quad (\text{por (4.3)}) \\ &\leq \sum_{n=0}^{\infty} |f_{n+1}|^2 \pi_{n+1} \quad (0 < p_n, q_{n+1} < 1) \\ &= \sum_{n=0}^{\infty} |f_n|^2 \pi_n - f_0 \pi_0 < \infty \end{aligned}$$

$$\begin{aligned} \sum_{n=0}^{\infty} |r_n f_n|^2 \pi_n &= \sum_{n=0}^{\infty} r_n^2 |f_n|^2 \pi_n \\ &\leq \sum_{n=0}^{\infty} |f_n|^2 \pi_n < \infty \quad (0 < r_n < 1) \end{aligned}$$

$$\begin{aligned} \sum_{n=0}^{\infty} |q_n f_{n-1}|^2 \pi_n &= \sum_{n=0}^{\infty} q_n^2 |f_{n-1}|^2 \pi_n \\ &= \sum_{n=0}^{\infty} q_n |f_{n-1}|^2 p_{n-1} \pi_{n-1} \quad (\text{por (4.3)}) \\ &\leq \sum_{n=0}^{\infty} |f_{n-1}|^2 \pi_{n-1} \quad (0 < p_{n-1}, q_n < 1) \\ &= \sum_{n=0}^{\infty} |f_n|^2 \pi_n < \infty \end{aligned}$$

Por tanto, estas tres sucesiones y también Pf son elementos de $\ell_\pi^2(\mathbb{N}_0)$.

- Lineal: Es trivial.
- Acotado: Probar que un operador lineal en un espacio de Hilbert está acotado es equivalente a probar que es continuo. Es decir, que existe una constante $K \geq 0$ tal

que $\|Pf\|_\pi \geq K\|f\|_\pi$. En efecto:

$$\begin{aligned}
\|Pf\|_\pi &= \left(\sum_{n=0}^{\infty} |p_n f_{n+1} + r_n f_n + q_n f_{n-1}|^2 \pi_n \right)^{1/2} \\
&\leq \left(\sum_{n=0}^{\infty} |p_n f_{n+1}|^2 \pi_n \right)^{1/2} + \left(\sum_{n=0}^{\infty} |r_n f_n|^2 \pi_n \right)^{1/2} + \left(\sum_{n=0}^{\infty} |q_n f_{n-1}|^2 \pi_n \right)^{1/2} \\
&= \left(\sum_{n=0}^{\infty} p_n^2 |f_{n+1}|^2 \pi_n \right)^{1/2} + \left(\sum_{n=0}^{\infty} r_n^2 |f_n|^2 \pi_n \right)^{1/2} + \left(\sum_{n=0}^{\infty} q_n^2 |f_{n-1}|^2 \pi_n \right)^{1/2} \\
&= \left(\sum_{n=0}^{\infty} p_n |f_{n+1}|^2 q_{n+1} \pi_{n+1} \right)^{1/2} + \left(\sum_{n=0}^{\infty} r_n^2 |f_n|^2 \pi_n \right)^{1/2} \\
&\quad + \left(\sum_{n=0}^{\infty} q_n |f_{n-1}|^2 p_{n-1} \pi_{n-1} \right)^{1/2} \\
&\leq \left(\sum_{n=0}^{\infty} |f_{n+1}|^2 \pi_{n+1} \right)^{1/2} + \left(\sum_{n=0}^{\infty} |f_n|^2 \pi_n \right)^{1/2} + \left(\sum_{n=0}^{\infty} |f_{n-1}|^2 \pi_{n-1} \right)^{1/2} \\
&\leq 3\|f\|_\pi,
\end{aligned}$$

donde hemos usado la desigualdad triangular, las ecuaciones de simetría (4.3) y que $0 < p_n, q_{n+1}, r_n, q_n, p_{n-1} < 1$. Por lo que el operador P es continuo y por tanto acotado en $\ell_\pi^2(\mathbb{N}_0)$.

- $\|P\| \leq 1$: Sabemos que $\|P\| = \min\{K \geq 0 : \|Pf\|_\pi^2 \leq \|f\|_\pi^2\}$. Sea $e = (1, 0, 0, \dots)$. Como $\|e\|_\pi = 1^2 \pi_0 = 1 < \infty$, tenemos que $e \in \ell_\pi^2(\mathbb{N}_0)$. Si aplicamos P a este elemento, tenemos que $Pe = (r_0, q_1, 0, 0, \dots)$, por lo que

$$\|Pe\|_\pi^2 = r_0^2 \pi_0 + q_1^2 \pi_1 = r_0^2 \pi_0 + q_1 p_0 \pi_0 \leq r_0 \pi_0 + p_0 \pi_0 = (r_0 + p_0) \pi_0 = 1.$$

Entonces $\|Pe\|_\pi = \|e\|_\pi$, de donde necesariamente $\|P\| \leq 1$.

- Autoadjunto: Recordemos la relación de simetría (4.3) ($\pi_n p_n = \pi_{n+1} q_{n+1}$). Entonces

$$\begin{aligned}
\langle Pf, g \rangle_\pi &= \sum_{n=0}^{\infty} (p_n f_{n+1} + r_n f_n + q_n f_{n-1}) g_n \pi_n \\
&= \sum_{n=0}^{\infty} p_n f_{n+1} g_n \pi_n + \sum_{n=0}^{\infty} r_n f_n g_n \pi_n + \sum_{n=0}^{\infty} q_n f_{n-1} g_n \pi_n \\
&= \sum_{n=0}^{\infty} q_{n+1} f_{n+1} g_n \pi_{n+1} + \sum_{n=0}^{\infty} r_n f_n g_n \pi_n + \sum_{n=0}^{\infty} p_{n-1} f_{n-1} g_n \pi_{n-1} \\
&= \sum_{n=0}^{\infty} q_n f_n g_{n-1} \pi_n + \sum_{n=0}^{\infty} r_n f_n g_n \pi_n + \sum_{n=0}^{\infty} p_n f_n g_{n+1} \pi_n \\
&= \sum_{n=0}^{\infty} f_n (p_n g_{n+1} + r_n g_n + q_n g_{n-1}) \pi_n = \langle f, Pg \rangle_\pi.
\end{aligned}$$

□

Con este lema nos aseguramos que el operador P verifica las condiciones del teorema espectral B.1 aunque, como se ha comentado con anterioridad, no es P el operador al que le aplicaremos el teorema espectral. Estudiaremos antes unos particulares elementos de $\ell_\pi^2(\mathbb{N}_0)$.

Para $i \geq 0$, denotamos con $e^{(i)} \in \ell_\pi^2(\mathbb{N}_0)$ a las sucesiones tales que $(e^{(i)})_j = \delta_{ij}/\pi_i$, que expresada en forma de vector columna tendrían la forma

$$e^{(i)} = (\overbrace{0, \dots, 0}^{(i-1)}, 1/\pi_i, 0, \dots)^T.$$

Si aplicamos el operador P a estos elementos, obtenemos

$$\begin{aligned} Pe^{(i)} &= (\overbrace{0, \dots, 0}^{(i-2)}, p_{i-1}/\pi_i, r_i/\pi_i, q_{i+1}/\pi_i, 0, \dots)^T \\ &= (0, \dots, 0, q_i/\pi_{i-1}, r_i/\pi_i, p_i/\pi_{i+1}, 0, \dots)^T \quad (\text{por (4.3)}) \\ &= q_i e^{(i-1)} + r_i e^{(i)} + p_i e^{(i+1)}. \end{aligned}$$

Hemos deducido la siguiente ecuación, que nos será útil próximamente en la demostración de la proposición 4.1:

$$Pe^{(n)} = q_n e^{(n-1)} + r_n e^{(n)} + p_n e^{(n+1)} \quad (n \geq 1) \quad (4.6)$$

Seguidamente introduciremos un resultado para el cual conviene recordar que, dado un polinomio cualquiera $\pi(x) = a_0 + a_1x + \dots + a_nx^n$ y una matriz cuadrada A , es posible evaluar el polinomio en A sin más que considerar

$$\pi(A) = a_0I + a_1A + \dots + a_nA^n.$$

Esta posibilidad nos ofrece una llamativa proposición.

Proposición 4.1. Para $n \geq 0$,

$$Q_n(P)e^{(0)} = e^{(n)} \quad (4.7)$$

Demostración. Lo probaremos por inducción.

- Caso base $n = 0$. Tenemos que $Q_0(x) = 1$, luego $Q_0(P) = I$. Por tanto $Q_0(P)e^{(0)} = e^{(0)}$.
- Caso base $n = 1$. En este caso, por la segunda ecuación de (4.4) y teniendo en cuenta que $Q_0(x) = 1$,

$$Q_1(x) = \frac{1}{p_0}(x - r_0).$$

Por tanto, $Q_1(P) = \frac{1}{p_0}(P - r_0I)$. Si multiplicamos por $e^{(0)} = (1, 0, \dots)^T$ obtenemos:

$$\begin{aligned} Q_1(P)e^{(0)} &= \frac{1}{p_0}(P - r_0I)e^{(0)} \\ &= \frac{1}{p_0} \begin{pmatrix} 0 & p_0 & 0 & \cdots & 0 \\ q_1 & r_1 - r_0 & p_1 & \cdots & 0 \\ 0 & q_2 & r_2 - r_0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{pmatrix} = \begin{pmatrix} 0 \\ q_1/p_0 \\ 0 \\ \vdots \end{pmatrix} = e^{(1)} \end{aligned}$$

- Inducción. Supongamos que la fórmula es cierta para valores $\leq n$ y comprobemos si es cierta para $n + 1$. Previamente, por la tercera ecuación de (4.4), obtenemos que

$$Q_{n+1}(x) = \frac{1}{p_n}((x - r_n)Q_n(x) - q_n Q_{n-1}(x)),$$

y si evaluamos matricialmente el polinomio en P y multiplicamos por $e^{(0)}$:

$$\begin{aligned} Q_{n+1}(P)e^{(0)} &= \frac{1}{p_n}((P - r_n I)Q_n(P)e^{(0)} - q_n Q_{n-1}(P)e^{(0)}) \\ &= \frac{1}{p_n}((P - r_n I)e^{(n)} - q_n e^{(n-1)}) \quad (\text{aplicando la hipótesis de inducción}) \\ &= \frac{1}{p_n}(Pe^{(n)} - r_n e^{(n)} - q_n e^{(n-1)}) \\ &= e^{(n+1)} \quad (\text{por la igualdad (4.6)}). \end{aligned}$$

□

A nivel intuitivo, esta proposición nos dice que los elementos $e^{(n)}$ tal como los hemos definido son la primera columna de la matriz infinita que resulta al evaluar el n -ésimo polinomio $Q_n(x)$ definido en (4.4) matricialmente en la matriz P definida en (4.1).

Seguidamente, buscaremos expresar la probabilidad $P_{ij}^{(n)}$ en función de productos escalares a los que podamos aplicar el teorema espectral. Así, tenemos

$$\begin{aligned} \langle P^n e^{(j)}, e^{(i)} \rangle_\pi &= \left\langle \left(\frac{1}{\pi_j} (P_{0,j}^n, P_{1,j}^n, \dots) \right)^T, \left(0, 0, \dots, \frac{1}{\pi_i}, \dots \right)^T \right\rangle_\pi \\ &= \frac{1}{\pi_j} \sum_{k=0}^{\infty} P_{k,j}^n \frac{\delta_{ik}}{\pi_i} \pi_k = \frac{P_{ij}^n}{\pi_j} = \frac{P_{ij}^{(n)}}{\pi_j}. \end{aligned}$$

Y si despejamos $P_{ij}^{(n)}$, obtenemos

$$P_{ij}^{(n)} = \pi_j \langle P^n e^{(j)}, e^{(i)} \rangle_\pi,$$

que aplicando la proposición 4.1 podemos escribir como

$$P_{ij}^{(n)} = \pi_j \langle P^n Q_j(P)e^{(0)}, Q_i(P)e^{(0)} \rangle_\pi.$$

Aplicando que P , y por tanto P^k para $k = 0, \dots, i$ es un operador autoadjunto, tenemos

$$P_{ij}^{(n)} = \pi_j \langle P^n Q_j(P)Q_i(P)e^{(0)}, e^{(0)} \rangle_\pi.$$

Atendiendo al último párrafo del apéndice B y aplicando el teorema espectral B.1, tomando $\mathcal{H} = \ell_\pi^2(\mathbb{N}_0)$, $f(x) = x^n Q_j(x)Q_i(x)$, $T = f(P)$, $x = y = e^{(0)}$, podemos afirmar que existe una medida $\psi = E_{e^{(0)}, e^{(0)}}$, cuyo soporte está contenido en $[-\|P\|, \|P\|] \subseteq [-1, 1]$, tal que

$$\int_{-1}^1 f(x) d\psi(x) = \langle f(P)e^{(0)}, e^{(0)} \rangle_\pi,$$

es decir

$$\int_{-1}^1 x^n Q_j(x)Q_i(x) d\psi(x) = \langle P^n Q_j(P)Q_i(P)e^{(0)}, e^{(0)} \rangle_\pi = \frac{P_{ij}^{(n)}}{\pi_j}.$$

De donde deducimos la fórmula de representación

$$P_{ij}^{(n)} = \pi_j \int_{-1}^1 x^n Q_j(x) Q_i(x) d\psi(x).$$

Resumimos esta deducción en el siguiente teorema.

Teorema 4.1 (Fórmula de Representación de Karlin-McGregor). *Sea $\{X_n, n \geq 0\}$ un proceso de nacimiento y muerte discreto con matriz de transición P definida como en (4.1). Entonces existe una única medida ψ , con soporte contenido en el intervalo $[-1, 1]$ tal que para cualesquiera estados i, j y cualquier instante n se tiene que*

$$P_{ij}^{(n)} = \pi_j \int_{-1}^1 x^n Q_j(x) Q_i(x) d\psi(x). \quad (4.8)$$

Demostración. La existencia está probada, la unicidad puede probarse teniendo en cuenta que tomando $i = j = 0$ se obtienen los momentos de ψ : $\int_{-1}^1 x^n d\psi(x)$. Estos momentos determinan unívocamente a la medida ψ . Consúltese [KM59, Theorem 1]. \square

REVIEW He puesto que la medida se denota con ψ mientras que en el primer capítulo puse μ , creo que convendría poner ambas iguales.

Como ya hemos adelantado anteriormente, de la medida ψ que nos proporciona el teorema espectral obtendremos la ortogonalidad de los polinomios $Q_n(x)$ dados en (4.4). Esta propiedad es, de hecho, una sencilla consecuencia de la fórmula de representación.

Corolario 4.2. *Los polinomios $\{Q_n\}$ definidos en (4.4) son ortogonales respecto a la medida ψ .*

Demostración. Si tomamos $n = 0$ en la fórmula de representación (4.8) y recordamos que $P^{(0)} = I$, la matriz identidad, que en este caso sería infinita, obtenemos

$$\int_{-1}^1 Q_j(x) Q_i(x) d\psi(x) = \frac{1}{\pi_j} \delta_{ij}.$$

\square

A la medida ψ se le suele denominar **medida espectral** y a los polinomios $\{Q_n\}$ **polinomios de nacimiento y muerte**.

A modo de reflexión final de esta sección y previo a mostrar un ejemplo clarificador, hasta el momento hemos probado que cualquier proceso de nacimiento y muerte admite una única medida espectral cuyo soporte está contenido en el intervalo $[-1, 1]$ y tal que sus polinomios de nacimiento y muerte, definidos en (4.4), forman una SPO. Además, las probabilidades de transición pueden ser calculadas utilizando integrales mediante la fórmula de representación de Karlin-McGregor (4.8). Sin embargo, es posible también estudiar el problema inverso, es decir, a partir de una medida ψ con soporte contenido en $[-1, 1]$, conocer la existencia y/o unicidad de los posibles procesos de nacimiento y muerte para los cuales ψ es su medida espectral. Por extensión no abordamos este aspecto en este trabajo, pero aconsejamos al lector interesado profundizar en [Dom22, Section 2.2] y [KM59].

Para finalizar esta sección, mostraremos un ejemplo en el que nos reencontraremos con los polinomios de Jacobi, introducidos en la sección 2.4.3. Este ejemplo está basado en la colección recogida en [Dom22, CH. II, Section 2.4], concretamente en la subsección 2.4.3.

Ejemplo 12 (Proceso de nacimiento y muerte generado por los polinomios de Jacobi). Consideremos el proceso de nacimiento y muerte dado por los parámetros

$$p_n = \frac{n + 2\alpha + 1}{2n + 2\alpha + 1} \quad r_n = 0 \quad q_n = \frac{n}{2n + 2\alpha + 1},$$

con $\alpha > -1$. Trataremos de calcular los polinomios de nacimiento y muerte, la medida espectral y la fórmula de representación. Comprobaremos que los polinomios de nacimiento y muerte $Q_n(x)$ son, en esencia, los polinomios de Jacobi presentados en la sección 2.4.3, tomando $\alpha = \beta$.

De acuerdo a las ecuaciones (4.4), los polinomios $Q_n(x)$ del proceso obedecen a la relación de recurrencia:

$$xQ_n(x) = \frac{n + 2\alpha + 1}{2n + 2\alpha + 1}Q_{n+1} + \frac{n}{2n + 2\alpha + 1}Q_{n-1}. \quad (4.9)$$

Los polinomios de Jacobi $P_n^{(\alpha, \beta)}(x)$, particularizando al caso $\alpha = \beta$, siguen la relación de recurrencia dada por (consúltese apéndice A):

$$(2n+2\alpha)_3 xP_n^{(\alpha, \alpha)}(x) = 2(n+1)(n+2\alpha+1)(2n+2\alpha)P_{n+1}^{(\alpha, \alpha)}(x) + 2(n+\alpha)^2(2n+2\alpha+2)P_{n-1}^{(\alpha, \alpha)}(x)$$

o equivalentemente

$$xP_n^{(\alpha, \alpha)}(x) = \frac{2(n+1)(n+2\alpha+1)}{(2n+2\alpha+1)(2n+2\alpha+2)}P_{n+1}^{(\alpha, \alpha)}(x) + \frac{n+\alpha}{2n+2\alpha+1}P_{n-1}^{(\alpha, \alpha)}(x) \quad (4.10)$$

Consideramos Q_n como la estandarización de los polinomios de Jacobi según:

$$Q_n(x) = \frac{P_n^{(\alpha, \alpha)}(x)}{P_n^{(\alpha, \alpha)}(1)}. \quad (4.11)$$

Si sustituimos en la ecuación (4.9) teniendo en cuenta que $P_n^{(\alpha, \alpha)}(1) = \frac{(\alpha+1)_n}{n!}$ (véase [AS64, Table 22.4]) obtenemos

$$xP_n^{(\alpha, \alpha)}(x) \frac{n!}{(\alpha+1)_n} = \frac{n+2\alpha+1}{2n+2\alpha+1} \frac{(n+1)!}{(\alpha+1)_{n+1}} P_{n+1}^{(\alpha, \alpha)}(x) + \frac{n}{2n+2\alpha+1} \frac{(n-1)!}{(\alpha+1)_{n-1}} P_{n-1}^{(\alpha, \alpha)}(x),$$

que mediante sencillas simplificaciones podemos comprobar que, en efecto, es equivalente a (4.10).

Por tanto, los polinomios de nacimiento y muerte de este proceso son los definidos en (4.11). Tenemos entonces que los polinomios $\{Q_n(x)\}$ son ortogonales en el intervalo $[-1, 1]$ y con respecto a la función peso

$$\rho(x) = c(1-x^2)^\alpha,$$

donde c es una constante de estandarización.

Por otro lado, mediante un sencillo cálculo tenemos

$$\begin{aligned}\pi_i &= \frac{p_0 \cdots p_{i-1}}{q_1 \cdots q_i} = \frac{1+2\alpha+1}{2+2\alpha+1} \cdot \frac{2+2\alpha+1}{4+2\alpha+1} \cdots \frac{i+2\alpha}{2(i-1)+2\alpha+1} \\ &= \frac{1}{2+2\alpha+1} \cdot \frac{2}{4+2\alpha+1} \cdots \frac{i}{2i+2\alpha+1} \\ &= \frac{(2\alpha+2)_{i-1}(2i+2\alpha+1)}{i!}.\end{aligned}\tag{4.12}$$

Finalmente, podemos aplicar la fórmula de representación (4.8) para obtener

$$\begin{aligned}P_{ij}^{(n)} &= P[X_n = j | X_0 = i] \\ &= \pi_j \int_{-1}^1 x^n Q_i(x) Q_j(x) \rho(x) dx \\ &= \frac{(2\alpha+2)_{j-1}(2j+2\alpha+1)}{j!} \int_{-1}^1 x^n \left[\frac{P_i^{(\alpha,\alpha)}(x)}{P_i^{(\alpha,\alpha)}(1)} \right] \left[\frac{P_j^{(\alpha,\alpha)}(x)}{P_j^{(\alpha,\alpha)}(1)} \right] \rho(x) dx \\ &= \frac{(2\alpha+2)_{j-1}(2j+2\alpha+1)}{(j!) P_i^{(\alpha,\alpha)}(1) P_j^{(\alpha,\alpha)}(1)} \int_{-1}^1 x^n P_i^{(\alpha,\alpha)}(x) P_j^{(\alpha,\alpha)}(x) \rho(x) dx.\end{aligned}$$

4.3. Análogos continuos de la fórmula de representación

Por ahora, en este capítulo y los anteriores hemos descrito la relación entre las cadenas de Markov, concretamente las cadenas de nacimiento y muerte, y los polinomios ortogonales. Esta relación se encuentra en la fórmula de representación de Samuel Karlin y James McGregor (4.8), mediante la cual podemos calcular fácilmente probabilidades mediante integrales de manera sencilla teniendo en cuenta la ortogonalidad de los polinomios $Q_n(x)$ descritos en (4.4).

Sin embargo, en todo momento nos hemos restringido al caso estrictamente discreto: cadenas de Markov discretas, procesos de nacimiento y muerte discretos... que nos llevan a una versión discreta de la fórmula de representación. Sin embargo, existe una definición continua de las cadenas de Markov y de los procesos de nacimiento y muerte a partir de la cual se obtiene una expresión continua de la fórmula de representación de Karlin McGregor.

En esta sección introduciremos superficialmente los análogos continuos de los procesos de nacimiento y muerte y la fórmula de representación de Karlin-McGregor. El siguiente texto está basado en las referencias [And91], [Sch00, Ch. 2, 3], [KM57] y [Dom22, Ch. III, Sections 3.1, 3.2], a las cuales remitimos al lector que desee profundizar.

4.3.1. Procesos de Markov continuos

Hasta el momento, hemos considerado el paso del tiempo como incrementos del valor de un parámetro $n = 0, 1, 2, \dots$ que representa el instante n -ésimo. En el caso continuo, de forma general consideramos que un instante viene dado por un valor $t \in [0, +\infty)$. Por su parte, el espacio de estados S no cambia: $S = \{0, 1, \dots, N\}$ en el caso finito o $S = \mathbb{N}_0$ en caso de que exista una cantidad infinita de posibles estados, que es el caso de los procesos

de nacimiento y muerte. Así, fijado $t \in [0, +\infty)$, llamamos $X_t \in S$ a la variable aleatoria “estado del sistema en el instante t ” y la colección de variables aleatorias $\{X_t, t \geq 0\}$ es un **proceso de Markov continuo**.

En este caso, la propiedad de Markov viene dada como:

$$P[X_{t+s} = j | X_s = i, X_\tau, 0 \leq \tau < s] = P[X_{t+s} = j | X_s = i], \quad s, t > 0. \quad (4.13)$$

Es decir, el estado en el que se encuentre el proceso en un instante $t+s$ depende únicamente del estado en el que se encuentre en un instante anterior t , independientemente de lo que haya ocurrido antes de t . De nuevo, nos centraremos en procesos homogéneos, que son aquellos en los cuales las probabilidades únicamente dependen de la diferencia $t - s$, por lo que podemos eliminar la dependencia de s y denotar

$$P_{ij}(t) := P[X_t = j | X_0 = i] \quad t > 0. \quad (4.14)$$

De esta forma, vemos que las probabilidades de transición se expresan en función del tiempo, de forma que lo que en el caso discreto era la matriz de transición P dada en (3.3), ahora es una función matricial $P(t)$:

$$P(t) = \begin{pmatrix} P_{00}(t) & P_{01}(t) & P_{02}(t) & \cdots \\ P_{10}(t) & P_{11}(t) & P_{12}(t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (4.15)$$

4.3.2. Procesos de nacimiento y muerte continuos

Particularizaremos al caso de procesos de nacimiento y muerte, en los que recordamos que su principal particularidad era la posibilidad de que en cualquier instante tan sólo era posible cambiar al siguiente estado, al anterior, o mantenerse en el mismo estado. Esta propiedad, llevada al modelo continuo, implica que si en un instante t el sistema se sitúa en el estado i , únicamente podemos permitir que $P_{i,i-1}(t)$, $P_{i,i}(t)$ o $P_{i,i+1}(t)$ sean no nulas. Por tanto, la matriz (4.15), que extiende a la matriz (4.1) al caso continuo, sería la siguiente:

$$P(t) = \begin{pmatrix} P_{00}(t) & P_{01}(t) & 0 & 0 & \cdots \\ P_{10}(t) & P_{11}(t) & P_{12}(t) & 0 & \cdots \\ 0 & P_{21}(t) & P_{22}(t) & P_{23}(t) & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}. \quad (4.16)$$

Además, asumiremos que las probabilidades (4.14) satisfacen las siguientes propiedades:

TODO Revisar cuando he puesto que los estados empiezan en 0 y cuando en 1 que tengo un follón montado curioso.

$$1. P_{i,i+1}(h) = \lambda_i h + o(h) \text{ cuando } h \searrow 0, i \geq 0.$$

$$2. P_{i,i-1}(h) = \mu_i h + o(h) \text{ cuando } h \searrow 0, i \geq 1.$$

3. $P_{i,i}(h) = 1 - (\lambda_i + \mu_i)h + o(h)$ cuando $h \searrow 0$, $i \geq 0$ asumiendo $\mu_0 = 0$ salvo que se indique lo contrario.
4. $P_{ij}(0) = \delta_{ij}$.
5. Asumimos que $\lambda_0 > 0$, $\lambda_i, \mu_i > 0$, $i \geq 1$.

Los parámetros λ_i, μ_i son las llamadas **tasas de nacimiento y muerte**, respectivamente. En los puntos 1 y 2 afirmamos que, si el proceso empieza en el estado i , en un intervalo pequeño de tiempo las probabilidades de avanzar o retroceder son aproximadamente proporcionales a la longitud del intervalo, siendo λ_i y μ_i las constantes que marcan esta proporcionalidad. En el punto 3 asumimos que no existen otras posibles transiciones que avanzar, retroceder o mantenerse en el estado i .

Las constantes $\{\mu_i\}, \{\lambda_i\}$ determinan por completo el comportamiento del proceso. Recogemos toda esta información en una matriz concreta.

Definición 4.2 (Matriz generadora infinitesimal). Dado un proceso de nacimiento y muerte continuo $\{X_t, t \geq 0\}$ caracterizado por las constantes $\{\mu_i\}, \{\lambda_i\}$, definimos su **matriz generadora infinitesimal** como:

$$\mathcal{A} = \begin{pmatrix} -(\lambda_0 + \mu_0) & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & \cdots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}. \quad (4.17)$$

Más información acerca de la génesis y la importancia de esta matriz en [Sch00, Section 2.1].

De forma similar al caso de procesos de nacimiento y muerte discretos, podemos definir a partir de $\{\mu_i\}$ y $\{\lambda_i\}$ los coeficientes potenciales (véase definición 4.1)

$$\pi_0 = 1 \quad \pi_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \quad (n \geq 1). \quad (4.18)$$

Estos coeficientes verifican las ecuaciones de simetría

$$\pi_n \lambda_n = \pi_{n+1} \mu_{n+1} \quad (n \geq 0). \quad (4.19)$$

La matriz (4.17) y los coeficientes (4.18) jugarán un papel esencial en el análogo continuo de la fórmula de representación.

4.3.3. La fórmula de representación de Karlin y McGregor continua

De manera análoga a la sección 4.2, presentaremos los llamados polinomios de nacimiento y muerte y la versión continua de la fórmula de representación.

Para empezar, a partir de la matriz generadora infinitesimal (4.17), definimos una sucesión de polinomios $\{Q_n\}$ con $\deg(Q_n) = n$:

$$\begin{aligned} Q_0(x) &= 1 \\ -xQ_0(x) &= \lambda_0 Q_1(x) - (\lambda_0 + \mu_0)Q_0(x), \\ -xQ_n(x) &= \lambda_n Q_{n+1}(x) - (\lambda_n + \mu_n)Q_n(x) + \mu_n Q_{n-1}(x), \quad (n \geq 1). \end{aligned} \quad (4.20)$$

De forma vectorial, las ecuaciones pueden expresarse como $-xQ(x) = \mathcal{A}Q(x)$, donde $Q(x) = (Q_0(x), Q_1(x), \dots)^T$. Estos polinomios son los que nos permiten obtener la fórmula de representación de Karlin-McGregor continua:

$$P_{ij}(t) = \pi_j \int_0^\infty e^{-xt} Q_j(x) Q_i(x) d\psi(x), \quad i, j \in \mathbb{N}_0, t \geq 0, \quad (4.21)$$

donde ψ es una medida cuyo soporte está contenido en el intervalo $[0, \infty)$ y para la cual los polinomios $\{Q_n\}$ definidos en (4.20) forman una SPO.

La demostración y los detalles de esta prueba pueden encontrarse en las referencias [KM57] o [Dom22, Section 3.2]. Finalmente, para clarificar el uso de esta fórmula, presentamos un ejemplo, basado en los trabajos sobre crecimiento lineal de Samuel Karlin y James McGregor [KM58].

Ejemplo 13 (Procesos de nacimiento y muerte lineales). Sea el proceso de nacimiento y muerte dado por los parámetros

$$\lambda_n = (n + \beta)\kappa \quad \mu_n = n\kappa,$$

donde $\beta, \kappa > 0$. Comprobaremos que los polinomios asociados a este proceso guardan una estrecha relación con los polinomios de Laguerre, introducidos en la sección 2.4.2.

Con estos parámetros, los polinomios de nacimiento y muerte $Q_n(x)$ siguen la relación de recurrencia

$$\begin{aligned} -xQ_n(x) &= \lambda_n Q_{n+1}(x) - (\lambda_n + \mu_n)Q_n(x) + \mu_n Q_{n-1}(x) \Leftrightarrow \\ -xQ_n(x) &= (n + \beta)\kappa Q_{n+1}(x) - (2n + \beta)\kappa Q_n(x) + n\kappa Q_{n-1}(x) \end{aligned}$$

Si dividimos la ecuación entre κ , obtenemos

$$-\frac{x}{\kappa}Q_n(x) = (n + \beta)Q_{n+1}(x) - (2n + \beta)Q_n(x) + nQ_{n-1}(x) \quad (4.22)$$

Los polinomios de Laguerre $L_n^\alpha(x)$ obedecen a una relación de recurrencia a tres términos dada por (consúltese apéndice A)

$$-xL_n^\alpha(x) = (n + 1)L_{n+1}^\alpha(x) - (2n + \alpha + 1)L_n^\alpha(x) + (n + \alpha)L_{n-1}^\alpha(x), \quad (4.23)$$

con $\alpha > -1$. Tomamos entonces

$$Q_n(x) = \frac{n!}{(\beta)_n} L_n^{\beta-1}\left(\frac{x}{\kappa}\right), \quad (4.24)$$

Si tomamos $\alpha = \beta - 1$, evaluamos (4.23) en x/κ y multiplicamos ambos miembros de la igualdad por $\frac{n!}{(\beta)_n}$, podemos aplicar las siguientes igualdades en los términos en $n + 1$ y en $n - 1$:

$$(n + 1) \left[\frac{n!}{(\beta)_n} \right] L_{n+1}^{\beta-1}\left(\frac{x}{\kappa}\right) = (n + \beta) \frac{(n + 1)!}{(\beta)_{n+1}} L_{n+1}^{\beta-1}\left(\frac{x}{\kappa}\right) = (n + \beta)Q_{n+1}(x),$$

y

$$(\beta + n - 1) \left[\frac{n!}{(\beta)_n} \right] L_{n-1}^{\beta-1}\left(\frac{x}{\kappa}\right) = n \frac{(n - 1)!}{(\beta)_{n-1}} L_{n-1}^{\beta-1}\left(\frac{x}{\kappa}\right) = nQ_{n-1}(x).$$

Aplicando estas igualdades, tenemos que (4.23) es equivalente a (4.22).

Por tanto, los polinomios asociados al proceso de nacimiento y muerte de este ejemplo son los presentados en (4.24). Esto es, los polinomios $\{Q_n\}$ son ortogonales en $[0, +\infty)$ con respecto a la función peso

$$\rho(x) = ce^{-x/\kappa}x^{\beta-1},$$

donde c es una constante de estandarización.

Podemos así calcular las probabilidades de transición aplicando la ecuación (4.18) y teniendo en cuenta que

$$\pi_i = \frac{\lambda_0 \cdots \lambda_{i-1}}{\mu_1 \cdots \mu_i} = \frac{\beta(\beta+1) \cdots (\beta+i-1)}{i!} = \frac{(\beta)_i}{i!}.$$

Tenemos entonces

$$\begin{aligned} P_{ij}(t) &= P[X_t = j | X_0 = i] \\ &= \pi_j \int_0^\infty e^{-xt} Q_i(x) Q_j(x) \rho(x) dx \\ &= c \frac{(\beta)_j}{j!} \int_0^\infty e^{-xt} \left[\frac{i!}{(\beta)_i} L_i^{\beta-1} \left(\frac{x}{\kappa} \right) \right] \left[\frac{j!}{(\beta)_j} L_j^{\beta-1} \left(\frac{x}{\kappa} \right) \right] e^{-\frac{x}{\kappa}} x^{\beta-1} dx \\ &= c \frac{i!}{(\beta)_i} \int_0^\infty L_i^{\beta-1} \left(\frac{x}{\kappa} \right) L_j^{\beta-1} \left(\frac{x}{\kappa} \right) e^{-x(t+\frac{1}{\kappa})} x^{\beta-1} dx. \end{aligned}$$

En conclusión, una vez presentada la teoría de polinomios ortogonales y la de procesos y cadenas de Markov de manera independiente, aunque aparentemente no parecían guardar ninguna relación la una con la otra, hemos podido descubrir la relación entre una y otra disciplina. Esta relación se sitúa en los procesos de nacimiento y muerte, en la que aprovechamos los tres coeficientes que caracterizan cada estado del sistema para crear lo que será una sucesión de polinomios ortogonales. Por su parte, la fórmula de representación de Karlin-McGregor permite encontrar un punto en común entre el cálculo de probabilidades y el de integrales, que aprovechando la ortogonalidad de los polinomios puede simplificarse considerablemente.

REVIEW No sé si poner que esto realmente facilita los cálculos de probabilidades. Aparentemente el cálculo de integrales no se ve más sencillo que el de probabilidades, pero teniendo en cuenta que aquí la cadena es infinita igual sí que compensa.

CONCLUSIONES

Las conclusiones

- [Álv03] Renato Álvarez Nodarse. *Polinomios hipergeométricos y q-polinomios (Monografías del Seminario matematico)*. Ene. de 2003.
- [Álv06] Renato Álvarez Nodarse. “On characterizations of classical polynomials”. En: *Journal of Computational and Applied Mathematics* 196.1 (2006), págs. 320-337. ISSN: 0377-0427. DOI: <https://doi.org/10.1016/j.cam.2005.06.046>. URL: <https://www.sciencedirect.com/science/article/pii/S0377042705005662>.
- [And91] William J. Anderson. *Continuous-time Markov chains : an applications-oriented approach*. eng. 1st ed. 1991. Springer Series in Statistics. New York, New York State: Springer-Verlag, 1991. ISBN: 1-4612-3038-1.
- [AS64] Milton Abramowitz e Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. ninth Dover printing, tenth GPO printing. New York: Dover, 1964.
- [AWH13] George B. Arfken, Hans J. Weber y Frank E. Harris. “Chapter 12 - Further Topics in Analysis”. En: *Mathematical Methods for Physicists (Seventh Edition)*. Ed. por George B. Arfken, Hans J. Weber y Frank E. Harris. Seventh Edition. Boston: Academic Press, 2013, págs. 551-598. ISBN: 978-0-12-384654-9. DOI: <https://doi.org/10.1016/B978-0-12-384654-9.00012-8>. URL: <https://www.sciencedirect.com/science/article/pii/B9780123846549000128>.
- [Chi11] Theodore Chihara. *An Introduction to Orthogonal Polynomials*. Dover Publications, feb. de 2011.
- [Dom22] Manuel Domínguez de la Iglesia. *Orthogonal polynomials in the spectral analysis of Markov processes : birth-death models and diffusion*. Encyclopedia of mathematics and its applications. Cambridge: Cambridge University Press, 2022. ISBN: 9781316516553.
- [DX14] Charles Dunkl y Yuan Xu. *Orthogonal Polynomials of Several Variables: 155*. 2nd Revised ed. Cambridge University Press, ago. de 2014.
- [Gro80] Emil Grosswald. “Bessel Polynomials”. En: *Journal of Applied Mathematics and Mechanics* 60.10 (1980), págs. 543-543. DOI: <https://doi.org/10.1002/zamm.19800601018>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/zamm.19800601018>.

- [INE] INE - Instituto Nacional de Estadística. *INE. Instituto Nacional de Estadística*. URL: <https://www.ine.es/index.htm> (visitado 28-03-2023).
- [KM57] Samuel Karlin y James L. McGregor. "The differential equations of birth-and-death processes, and the Stieltjes moment problem". En: *Transactions of the American Mathematical Society* 85 (1957), págs. 489-546.
- [KM58] Samuel Karlin y James McGregor. "Linear Growth, Birth and Death Processes". En: *Journal of Mathematics and Mechanics* 7.4 (1958), págs. 643-662. ISSN: 00959057, 19435274. URL: <http://www.jstor.org/stable/24900526> (visitado 03-03-2023).
- [KM59] Samuel Karlin y James McGregor. "Random walks". En: *Illinois Journal of Mathematics* 3.1 (1959), págs. 66-81.
- [Kul12] V. Kulkarni. *Introduction to Modeling and Analysis of Stochastic Systems*. New York, Estados Unidos: Springer Publishing, 2012.
- [MBP94] F. Marcellán, A. Branquinho y J Petronilho. "Classical orthogonal polynomials: A functional approach". En: *Acta Applicandae Mathematica* 34 (1994), págs. 283-303. DOI: [10.1007/BF00998681](https://doi.org/10.1007/BF00998681). URL: <https://doi.org/10.1007/BF00998681>.
- [Moy09] Nancy Rosa Moya Lázaro. "Funciones de peso". En: *Pesquimat* 12.1 (jul. de 2009). DOI: [10.15381/pes.v12i1.9488](https://doi.org/10.15381/pes.v12i1.9488). URL: <https://revistasinvestigacion.unmsm.edu.pe/index.php/matema/article/view/9488>.
- [Pay20] Rafael Payá. *Apuntes de Análisis Funcional. Tema 3: Operadores y funcionales lineales continuos*. 2020. URL: https://www.ugr.es/~rpaya/documentos/Funcional/2020-21/Apuntes_03.pdf (visitado 13-06-2023).
- [Ros83] Sheldon M. Ross. *Stochastic processes*. eng. Wiley series in probability and mathematical statistics. Probability and mathematical statistics. New York: John Wiley, 1983. ISBN: 0471099422.
- [Rud91] W. Rudin. *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1991. ISBN: 9780070542365. URL: https://books.google.es/books?id=Sh%5C_vAAAAMAAJ.
- [Sch00] Wim Schoutens. *Stochastic Processes and Orthogonal Polynomials*. New York, Estados Unidos: Springer Publishing, 2000.
- [Sen81] E. Seneta. *Non-negative matrices and Markov chains*. eng. Second edition. Springer series in statistics. New York: Springer-Science+Business Media, LLC, 1981. ISBN: 1-280-61900-7.
- [Ste18] Stefan Steinerberger. *Electrostatic Interpretation of Zeros of Orthogonal Polynomials*. 2018. DOI: [10.48550/ARXIV.1804.09697](https://doi.org/10.48550/ARXIV.1804.09697). URL: <https://arxiv.org/abs/1804.09697> (visitado 07-03-2023).
- [Sze75] G. Szego. *Orthogonal Polynomials*. American Math. Soc: Colloquium publ. American Mathematical Society, 1975. ISBN: 9780821810231. URL: <https://books.google.es/books?id=Z0hmnsXlcYOC>.
- [Tea23a] The Sage Development Team. *2D Graphics. Release 9.8*. 2023. URL: <https://doc.sagemath.org/pdf/en/reference/plotting/plotting.pdf> (visitado 07-03-2023).

[Tea23b] The Sage Development Team. *Orthogonal polynomials - Functions*. 2023. URL: https://doc.sagemath.org/html/en/reference/functions/sage/functions/orthogonal_polys.html (visitado 07-03-2023).

APÉNDICE A

TABLAS SOBRE POLINOMIOS ORTOGONALES CLÁSICOS

En este apéndice puede consultarse todo lo explicado sobre los polinomios ortogonales clásicos reales: ecuación de Pearson, intervalo de ortogonalidad, función peso, RR3T, ecuación diferencial hipergeométrica, fórmula de Rodrigues... Apéndice basado en las deducciones del capítulo 2 y en las tablas de [AS64, Ch. 22].

Familia de polinomios	Notación	Parámetros	Intervalo de ortogonalidad I
Hermite	$H_n(x)$	-	$\mathbb{R} = (-\infty, +\infty)$
Laguerre	$L_n^{(\alpha)}(x)$	$\alpha > -1$	$\mathbb{R}_0^+ = [0, +\infty)$
Jacobi	$P_n^{(\alpha, \beta)}(x)$	$\alpha, \beta > -1$	$[-1, 1]$

Tabla A.1: Generalidades sobre los PO clásicos

A.1. Relación de recurrencia a tres términos

Los PO clásicos satisfacen la relación de recurrencia

$$a_{1n}xP_n(x) = a_{2n}P_{n+1}(x) - a_{3n}P_n(x) + a_{4n}P_{n-1}(x) \quad \forall n \in \mathbb{N}_0, \quad P_{-1} = 0, P_0 = 1,$$

donde

Familia	a_{1n}	a_{2n}	a_{3n}	a_{4n}
Hermite	2	1	0	$2n$
Laguerre	-1	$n + 1$	$2n + \alpha + 1$	$n + \alpha$
Jacobi	$(2n + \alpha + \beta)_3$	$\frac{2(n+1) \times}{\times (n + \alpha + \beta + 1) \times}$	$\frac{(2n + \alpha + \beta + 1) \times}{\times (\alpha^2 - \beta^2)}$	$\frac{2(n + \alpha)(n + \beta) \times}{\times (2n + \alpha + \beta + 2)}$

Tabla A.2: Parámetros de la RR3T de las SPO clásicas

A.2. Ecuación de Pearson

Las SPO clásicas son ortogonales respecto a una función peso $\rho(x)$ definida en I que es solución de la ecuación de Pearson

$$[\sigma(x)\rho(x)]' = \tau(x)\rho(x),$$

donde:

Familia	$\sigma(x)$	$\tau(x)$	$\rho(x)$
Hermite	1	$-2x$	e^{-x^2}
Laguerre	x	$-x + \alpha + 1$	$x^\alpha e^{-x}$
Jacobi	$1 - x^2$	$-(\alpha + \beta + 2)x + (\beta - \alpha)$	$(1 - x)^\alpha (1 + x)^\beta$

Tabla A.3: Parámetros de la ecuación de Pearson

A.3. Ecuación diferencial hipergeométrica

Los polinomios de Hermite, Laguerre y Jacobi son soluciones, para todo $n \in \mathbb{N}_0$, de las siguientes EDOs:

Familia	Ecuación diferencial hipergeométrica
Hermite	$(H_n)'' - 2x(H_n)' + 2nH_n$
Laguerre	$x(L_n^{(\alpha)})'' + [-x + \alpha + 1](L_n^{(\alpha)})' + nL_n^{(\alpha)}$
Jacobi	$[1 - x^2](P_n^{(\alpha, \beta)})'' + [-(\alpha + \beta + 2)x + (\beta - \alpha)](P_n^{(\alpha, \beta)})' + n(\alpha + \beta + n + 1)P_n^{(\alpha, \beta)}$

Tabla A.4: Ecuación diferencial que satisfacen los PO clásicos.

A.4. Fórmulas de Rodrigues

Los polinomios de Hermite, Laguerre y Jacobi pueden ser calculados mediante las siguientes fórmulas de Rodrigues:

Familia	Fórmula de Rodrigues
Hermite	$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}$
Laguerre	$L_n^{(\alpha)}(x) = \frac{1}{n!} x^{-\alpha} e^x \frac{d^n}{dx^n} (x^{n+\alpha} e^{-x})$
Jacobi	$P_n^{(\alpha, \beta)}(x) = \frac{(-1)^n}{2^n n!} (1 - x)^{-\alpha} (1 + x)^{-\beta} \frac{d^n}{dx^n} (1 - x)^{n+\alpha} (1 + x)^{n+\beta}$

Tabla A.5: Fórmulas de Rodrigues de los PO clásicos.

En este apéndice introducimos una herramienta fundamental en el camino en busca de la fórmula de representación de Karlin-McGregor: el teorema espectral. Gracias a este teorema obtendremos una medida que nos proporcionará la ortogonalidad de los polinomios. Estos conceptos, al no ser tan cercanos al objetivo de este trabajo, se introducen de manera superficial, recomendando al lector que si desea profundizar en ellos consulte bibliografía relacionada con análisis funcional. Por ejemplo, se propone [Dom22, Ch. 1] o [Rud91, Ch. 12].

[TODO Añadir definición de norma de operador](#)

Para empezar, recordamos algunos conceptos básicos de análisis funcional.

Definición B.1 (Espacio de Hilbert). Un espacio vectorial \mathcal{H} se denomina **espacio de Hilbert** si está dotado de un producto escalar interno $\langle \cdot, \cdot \rangle$ y es además completo respecto a la norma asociada a dicho producto escalar.

Denotamos como $\mathcal{B}(\mathcal{H})$ al conjunto de todos los operadores lineales de \mathcal{H} en sí mismo. Sea $T : \mathcal{H} \rightarrow \mathcal{H}$ un operador lineal. T es un operador acotado si, dado un subconjunto $A \subseteq \mathcal{H}$ acotado, $T(A)$ también es acotado. Es decir, un operador lineal acotado transforma conjuntos acotados en conjuntos acotados. Equivalentemente, un operador lineal es acotado si, y solo si es continuo, es decir, si existe una constante $K \geq 0$ tal que, para todo $x \in \mathcal{H}$, $\|Tx\| \leq K\|x\|$. A partir de esta equivalencia presentamos la siguiente definición.

Definición B.2. Sea \mathcal{H} un espacio de Hilbert y $T : \mathcal{H} \rightarrow \mathcal{H}$ un operador lineal continuo (\Leftrightarrow acotado). Se define la **norma del operador** T como

$$\|T\| := \min\{K \geq 0 : \|Tx\| \leq K\|x\|\}. \quad (\text{B.1})$$

Definimos una propiedad particular que pueden verificar los operadores lineales.

Definición B.3 (Operador autoadjunto). Sea \mathcal{H} un espacio de Hilbert. Un operador lineal $T : \mathcal{H} \rightarrow \mathcal{H}$ es **autoadjunto** si

$$\langle Tx, y \rangle = \langle x, Ty \rangle, \quad \forall x, y \in \mathcal{H}. \quad (\text{B.2})$$

Por último, recordemos que el espectro de una matriz cuadrada $A \in \mathbb{R}^{N \times N}$, denotado como $\sigma(A)$ es el conjunto de sus valores propios. Es decir, los números $\lambda \in \mathbb{C}$ tales que el determinante de la matriz $A - \lambda I_N$ se anulaba. Dicho de otra manera, si nos planteamos el endomorfismo de \mathbb{R}^N $x \mapsto xA$, $\sigma(A)$ es el conjunto de valores λ tales que el endomorfismo $x \mapsto x(A - \lambda I)$ no es invertible. A partir de esta idea, podemos generalizar el concepto de espectro a cualquier operador.

Definición B.4 (Conjuntos espectro y resolvente). Sea $T : \mathcal{H} \longrightarrow \mathcal{H}$ un operador lineal sobre un espacio de Hilbert.

- El **resolvente** de T , denotado como $\rho(T)$ es el conjunto de valores $\lambda \in \mathbb{C}$ tales que el operador $T - \lambda I$ (también denominado **operador resolvente**) es invertible.
- El **espectro** de T , denotado como $\sigma(T)$ es el conjunto de valores $\lambda \in \mathbb{C}$ tales que el operador resolvente $T - \lambda I$ no es invertible.

Naturalmente, $\sigma(T)$ y $\rho(T)$ son conjuntos complementarios.

Si un operador lineal $T \in \mathcal{B}(\mathcal{H})$ es acotado, entonces $\sigma(T)$ es un subconjunto compacto del disco de radio $\|T\|$. Si además T es autoadjunto, $\sigma(T) \subseteq \mathbb{R}$, por lo que $\sigma(T) \subseteq [-\|T\|, \|T\|]$.

Por último, un concepto que próximamente nos facilitará la medida de ortogonalidad.

Definición B.5 (Resolución de la identidad). Una **resolución de la identidad** E en el espacio de Hilbert \mathcal{H} es un operador $E : \mathbb{R} \longrightarrow \mathcal{B}(\mathcal{H})$ que, para cualesquiera subconjuntos Borel-medibles $A, B \subseteq \mathbb{R}$, verifica las siguientes condiciones:

1. $E(A)$ es una proyección, es decir $E(A) \circ E(A) = E(A)$.
2. $E(A \cap B) = E(A)E(B)$.
3. $E(\emptyset) = 0_{\mathcal{B}(\mathcal{H})}$, $E(\mathbb{R}) = I_{\mathcal{H}}$.
4. Si A y B son disjuntos, $E(A \cup B) = E(A) + E(B)$.
5. La aplicación $A \longmapsto E_{x,y}(A) = \langle E(A)x, y \rangle$ es una medida compleja para cualesquiera $x, y \in \mathcal{H}$.

A modo introductorio, el teorema espectral nos facilita una resolución de la identidad a partir de un operador lineal. Con todos estos ingredientes, podemos presentar el enunciado esperado.

Teorema B.1 (Teorema espectral). Sea \mathcal{H} un espacio de Hilbert y $T \in \mathcal{B}(\mathcal{H})$ un operador lineal acotado y autoadjunto. Entonces existe una única resolución de la identidad E en \mathcal{H} tal que $T = \int_{\mathbb{R}} t dE(t)$, i.e.

$$\langle Tx, y \rangle = \int_{\mathbb{R}} t dE_{x,y}(t). \quad (\text{B.3})$$

Además, el soporte de E está contenido en $\sigma(T) \subseteq [-\|T\|, \|T\|]$.

Demostración. Consúltese [Rud91, Theorem 12.22]. □

Como última apreciación, pensemos en una función continua f definida sobre el espectro $\sigma(T)$ de un operador. Podemos definir entonces el operador

$$f(T) = \int_{\mathbb{R}} f(t) dE(t),$$

de forma que

$$\langle f(T)x, y \rangle = \int_{\mathbb{R}} f(t) dE_{x,y}(t).$$

Entonces el operador $f(T)$ es acotado y tiene norma $\|f(T)\| = \sup_{x \in \sigma(T)} |f(x)|$.