1

Orthogonal Polynomials

In this chapter we introduce some basic definitions and properties about the theory of special functions and orthogonal polynomials on the real line. In the first section we will introduce some basic special functions and the concept of the *Stieltjes transform*, which will be used frequently in the text. In Section 1.2 we will give some properties of the general theory of orthogonal polynomials. Section 1.3 is devoted to the *spectral theorem* and in particular applied to orthogonal polynomials, in which case it is usually called *Favard's theorem*. In Sections 1.4 and 1.5 we will focus on the so-called *classical orthogonal polynomials*, both of a continuous and a discrete variable. These special families, apart from being orthogonal, are characterized by the fact that they are eigenfunctions of a second-order differential operator (in the continuous variable) or a second-order difference operator (in the discrete variable) of the *Sturm–Liouville* type. Finally, in Section 1.6, we describe the *Askey scheme*, which is a way of organizing orthogonal polynomials of hypergeometric type into a hierarchy, where the classical orthogonal polynomials are included. This chapter is based on references [3, 9, 16, 74, 135, 137, 142].

1.1 Some Special Functions and the Stieltjes Transform

The *Gamma function* is a complex-valued function that extends the domain of the factorial function of a nonnegative integer $n!$. It was introduced by Euler in 1789 and it is defined by its integral representation

$$\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt, \quad \operatorname{Re} z > 0. \tag{1.1}$$

Integrating by parts we obtain the functional equation

$$z\Gamma(z) = \Gamma(z+1), \quad \operatorname{Re} z > 0.$$

The formula above can also be written as

$$(z)_n \Gamma(z) = \Gamma(z+n), \quad n \geq 0,$$

where $(z)_n$ is the *Pochhammer symbol*

$$(z)_n = \begin{cases} 1, & \text{if } n = 0, \\ z(z+1)\cdots(z+n-1), & \text{if } n \geq 1. \end{cases} \tag{1.2}$$

From here we also observe that if n is a nonnegative integer, then $\Gamma(n+1) = n!$.

The *Beta function* is defined by the integral

$$\mathrm{B}(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt, \quad \mathrm{Re}\, x,\ \mathrm{Re}\, y > 0. \tag{1.3}$$

It is symmetric, i.e. $\mathrm{B}(x,y) = \mathrm{B}(y,x)$, and it is related to the Gamma function by the well-known formula

$$\mathrm{B}(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}.$$

A *hypergeometric series* $\sum_{n=0}^{\infty} c_n$ is a series for which $c_0 = 1$ and the ratio of consecutive terms is a rational function of the summation index n, i.e. one for which

$$\frac{c_{n+1}}{c_n} = \frac{P(n)}{Q(n)},$$

where $P(n)$ and $Q(n)$ are polynomials. In this case, c_n is called a hypergeometric term. If the polynomials are completely factored, the ratio of successive terms can be written as

$$\frac{c_{n+1}}{c_n} = \frac{P(n)}{Q(n)} = \frac{(n+a_1)(n+a_2)\cdots(n+a_p)}{(n+b_1)(n+b_2)\cdots(n+b_q)(n+1)},$$

where the factor $n+1$ in the denominator is present for historical reasons of notation. From here we define the *generalized hypergeometric function* as

$${}_pF_q\left(\begin{matrix} a_1, \ldots, a_p \\ b_1, \ldots, b_q \end{matrix}; x\right) = \sum_{n=0}^{\infty} c_n x^n = \sum_{n=0}^{\infty} \frac{(a_1)_n \cdots (a_p)_n}{(b_1)_n \cdots (b_q)_n} \frac{x^n}{n!}. \tag{1.4}$$

We can also use the following notation for generalized hypergeometric functions:

$${}_pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; x).$$

This series is absolutely convergent for all x if $p \leq q$ and for $|x| < 1$ if $p = q+1$. It is divergent for all $x \neq 0$ if $p > q+1$, as long as the series is not finite. Observe that when one of the parameters of the numerator $a_i, i = 1, \ldots, p$, is a negative integer, then the generalized hypergeometric function is a polynomial.

Many of the known special functions can be represented in terms of generalized hypergeometric functions. For example, the simplest cases of ${}_0F_0$ and ${}_1F_0$ correspond to the exponential series and the binomial series, respectively. Indeed,

$${}_0F_0(-;\, -\,; x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x,$$

$$_1F_0(a;\,-\,;x)=\sum_{n=0}^{\infty}\frac{(a)_n x^n}{n!}=\sum_{n=0}^{\infty}\frac{\Gamma(z+n)}{\Gamma(a)\Gamma(n+1)}x^n=\sum_{n=0}^{\infty}\binom{a+n-1}{n}x^n=(1-x)^{-a}.$$

If $p=2$ and $q=1$, the function becomes what is called the *Gaussian hypergeometric function* $_2F_1(a,b;c;x)$ and it is related to the solutions of *Euler's hypergeometric differential equation*

$$x(1-x)y''(x)+[c-(a+b+1)x]y'(x)-aby(x)=0. \tag{1.5}$$

We will see later the relation of this equation with the Jacobi polynomials. All families of orthogonal polynomials in the Askey scheme admit a representation in terms of hypergeometric series, as we will see later. For more information about generalized hypergeometric functions see [3, Chapter 2].

The *Stieltjes transform* (also known as the Cauchy transform) of a measure ψ defined on $\mathbb{R}$ is defined as the complex-valued function

$$B(z;\psi)=\int_{\mathbb{R}}\frac{d\psi(x)}{x-z},\quad z\in\mathbb{C}\setminus\mathbb{R}. \tag{1.6}$$

This transform is related to the *generating function of the moments* of the measure ψ, since, formally

$$B(z;\psi)=-\frac{1}{z}\int_{\mathbb{R}}\frac{1}{1-x/z}d\psi(x)=-\frac{1}{z}\sum_{n=0}^{\infty}\int_{\mathbb{R}}\frac{x^n}{z^n}\,d\psi(x)=-\sum_{n=0}^{\infty}\frac{\mu_n}{z^{n+1}}, \tag{1.7}$$

where $\mu_n=\int_{\mathbb{R}}x^n d\psi(x)$ are the *moments* of the measure. In the case where $\operatorname{supp}(\psi)\subseteq[-A,A]$, then $|\mu_n|\leq 2A^n$, implying that the series (1.7) is absolutely convergent for $|z|>A$. In this case, the Stieltjes transform is completely determined in terms of the moments of the measure ψ. In general, the expansion of the Stieltjes transform (1.6) has to be interpreted as an asymptotic expansion of the Stieltjes transform $B(z;\psi)$ as $|z|\to\infty$.

There is a formula which allows to calculate the measure ψ if we have information about the corresponding Stieltjes transform. This formula is known as the *Perron–Stieltjes inversion formula*. It has several versions, but the one we will use in this text is included in the following result.

Proposition 1.1 ([51, Theorem X.6.1]) *Let ψ be a probability measure with finite moments and $B(z;\psi)$ its Stieltjes transform* (1.6). *Then*

$$\int_a^b d\psi(x)+\frac{1}{2}\psi(\{a\})+\frac{1}{2}\psi(\{b\})=\frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\int_a^b \operatorname{Im}B(x+i\varepsilon;\psi)\,dx, \tag{1.8}$$

where $\psi(\{a\})\geq 0$ is the magnitude or size of the mass at an isolated point a. If the measure is absolutely continuous at a then $\psi(\{a\})=0$.

Proof Observe that

$$2i\mathrm{Im}B(z;\psi) = B(z;\psi) - \overline{B(z;\psi)} = B(z;\psi) - B(\bar{z};\psi) = \int_{\mathbb{R}} \left[\frac{1}{x-z} - \frac{1}{x-\bar{z}}\right] d\psi(x)$$
$$= \int_{\mathbb{R}} \frac{z-\bar{z}}{|x-z|^2} d\psi(x) = 2i \int_{\mathbb{R}} \frac{\mathrm{Im}z}{|x-z|^2} d\psi(x).$$

Therefore

$$\mathrm{Im}B(x+i\varepsilon;\psi) = \int_{\mathbb{R}} \frac{\varepsilon}{|s-(x+i\varepsilon)|^2} d\psi(s) = \int_{\mathbb{R}} \frac{\varepsilon}{(s-x)^2+\varepsilon^2} d\psi(s).$$

Integrating and exchanging integrals (which is allowed since the integrand is positive) we have that

$$\int_a^b \mathrm{Im}B(x+i\varepsilon;\psi)\,dx = \int_{\mathbb{R}} \left[\int_a^b \frac{\varepsilon}{(s-x)^2+\varepsilon^2} dx\right] d\psi(s).$$

The internal integral can be calculated explicitly by making the change of variables $y = (x-s)/\varepsilon$:

$$\chi_\varepsilon(s) = \int_a^b \frac{\varepsilon}{(s-x)^2+\varepsilon^2} dx = \int_{(a-s)/\varepsilon}^{(b-s)/\varepsilon} \frac{1}{1+y^2} dy = \arctan y \Big|_{y=(a-s)/\varepsilon}^{y=(b-s)/\varepsilon}.$$

We have that $0 \leq \chi_\varepsilon(s) \leq \pi$ and when we take the limit (which is also allowed using the Lebesgue dominated convergence theorem since ψ is a probability measure and $\chi_\varepsilon(s)$ is bounded and positive) we have that

$$\lim_{\varepsilon\downarrow 0} \chi_\varepsilon(s) = \begin{cases} \pi, & \text{if} \quad a < s < b, \\ \frac{\pi}{2}, & \text{if} \quad s = a \text{ or } s = b. \end{cases}$$ □

As a consequence of the previous proposition we also have the formula

$$\int_a^b d\psi(x) = \frac{1}{\pi} \lim_{\varepsilon\downarrow 0} \lim_{\eta\downarrow 0} \int_{a+\eta}^{b-\eta} \mathrm{Im}B(x+i\varepsilon;\psi)\,dx. \tag{1.9}$$

When the measure is absolutely continuous with respect to the Lebesgue measure, i.e. $d\psi(x) = \psi(x)\,dx$ (abusing the notation), we have

$$\psi(x) = \frac{1}{\pi} \lim_{\varepsilon\downarrow 0} \mathrm{Im}B(x+i\varepsilon;\psi) = \lim_{\varepsilon\downarrow 0} \frac{B(x+i\varepsilon;\psi) - B(x-i\varepsilon;\psi)}{2\pi i}. \tag{1.10}$$

Finally, for measures that have an absolutely continuous part and a discrete part, there is a direct way to calculate the size of the jump. Indeed, assume that $\psi = \widehat{\psi} + \psi(\{a\})\delta_a$, where $\delta_a(x) = \delta(x-a)$ is the Dirac delta distribution which is defined, as usual, by $\int_{\mathbb{R}} f(x)\delta(x-a)\,dx = f(a)$. Then, since the Stieltjes transform is linear, we have

$$B(z;\psi) = B(z;\widehat{\psi}) + \frac{\psi(\{a\})}{a-z}.$$

Evaluating at $z = a + i\varepsilon$ and taking imaginary parts, we have

$$\mathrm{Im}B(a+i\varepsilon;\psi) = \mathrm{Im}B(a+i\varepsilon;\widehat{\psi}) + \mathrm{Im}\frac{\psi(\{a\})}{-i\varepsilon} = \mathrm{Im}B(a+i\varepsilon;\widehat{\psi}) + \frac{\psi(\{a\})}{\varepsilon}.$$

Therefore we get

$$\psi(\{a\}) = \varepsilon\mathrm{Im}B(a+i\varepsilon;\psi) - \varepsilon\mathrm{Im}B(a+i\varepsilon;\widehat{\psi}). \tag{1.11}$$

Taking limits as $\varepsilon \downarrow 0$ we observe that $B(a+i\varepsilon;\widehat{\psi})$ is bounded since $\widehat{\psi}$ is absolutely continuous. Therefore the meaningful isolated points (where $\psi(\{a\}) > 0$) must be those satisfying

$$\lim_{\varepsilon\downarrow 0}\mathrm{Im}B(a+i\varepsilon;\psi) = \infty,$$

while the size of the jump at $x = a$ is given by

$$\psi(\{a\}) = \lim_{\varepsilon\downarrow 0}\varepsilon\mathrm{Im}B(a+i\varepsilon;\psi) \geq 0. \tag{1.12}$$

Example 1.2 Let $B(z;\psi)$ be given by

$$B(z;\psi) = \frac{1}{1-z}, \quad z \in \mathbb{C}\setminus\{1\}.$$

According to (1.11) there will be a pole at $z = 1$, so it is a candidate for a singular part of the measure. Assume that $\psi = \widehat{\psi} + \psi(\{1\})\delta_1$, where $\widehat{\psi}$ is the absolutely continuous part. Then, by (1.10), we have

$$\widehat{\psi}(x) = \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\mathrm{Im}\frac{1}{1-x-i\varepsilon} = \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\mathrm{Im}\left(\frac{1-x+i\varepsilon}{(1-x)^2+\varepsilon^2}\right) = \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\frac{\varepsilon}{(1-x)^2+\varepsilon^2}.$$

We observe that if $x \neq 1$, then $\widehat{\psi}(x) = 0$. Therefore the measure ψ consists only of a singular part at $x = 1$. The value of $\psi(\{1\})$ is given by (1.12) and it is easy to see that

$$\psi(\{1\}) = \lim_{\varepsilon\downarrow 0}\varepsilon\mathrm{Im}B(1+i\varepsilon;\psi) = \lim_{\varepsilon\downarrow 0}\varepsilon\frac{\varepsilon}{\varepsilon^2} = 1.$$

Therefore $\psi(x) = \delta_1(x)$. ◊

Example 1.3 Consider the Stieltjes transform given by

$$B(z;\psi) = -2z + 2\sqrt{z^2-1}, \quad z \in \mathbb{C}\setminus[-1,1],$$

where the branch of the square root is determined by analytic continuation from positive values for real $z > 1$. We observe that there are no singular points, so the measure will consist only of an absolutely continuous part. From (1.10) we get

$$\begin{aligned}\psi(x) &= \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\mathrm{Im}B(x+i\varepsilon;\psi)\\ &= \frac{1}{\pi}\lim_{\varepsilon\downarrow 0}\left(-2\varepsilon + 2\mathrm{Im}\sqrt{x^2-\varepsilon^2+2ix\varepsilon-1}\right) = \frac{2}{\pi}\mathrm{Im}\sqrt{x^2-1}.\end{aligned}$$

The last part has only imaginary part when $|x| \leq 1$. Therefore

$$\psi(x) = \frac{2}{\pi}\sqrt{1-x^2}, \quad |x| < 1,$$

which is the *Wigner semicircle distribution*. ◊

In Chapters 2 and 3 we will see several examples of computation of measures using the Perron–Stieltjes inversion formula.

Remark 1.4 As we have seen in (1.7), the Stieltjes transform is related to the generating function of the moments of a probability measure ψ. This is not exactly the same as the usual *moment generating function*, which is defined as

$$M_X(t) = \mathbb{E}(e^{tX}) = \sum_{n=0}^{\infty} \mu_n \frac{t^n}{n!},$$

where X is the random variable associated with the probability measure ψ. This moment generating function is more related to the *Laplace transform*. Indeed, assume that the probability measure is absolutely continuous and supported on $[0,\infty)$. Then the Laplace transform is defined by

$$\mathcal{L}[\psi](s) = \int_0^\infty e^{-sx}\psi(x)\,dx.$$

Then we have $\mathcal{L}[\psi](-t) = M_X(t)$. The Stieltjes transform arises naturally as an iteration of the Laplace transform. Indeed, if we call $\phi(s) = \mathcal{L}[\psi](s)$ then, formally, we have

$$\begin{aligned}
\mathcal{L}[\phi](t) &= \int_0^\infty e^{-st}\phi(s)ds = \int_0^\infty e^{-st}\left(\int_0^\infty e^{-su}\psi(u)du\right)ds \\
&= \int_0^\infty \psi(u)\left(\int_0^\infty e^{-s(t+u)}ds\right)du \\
&= \int_0^\infty \psi(u)\left(-\frac{1}{t+u}e^{-s(t+u)}\Big|_{s=0}^{s=\infty}\right)du = \int_0^\infty \frac{\psi(u)}{t+u}du, \quad \mathrm{Re}(t) > 0.
\end{aligned}$$

Therefore $B(t;\psi) = \mathcal{L}^2[\psi](-t)$. A good reference about Stieltjes transforms in connection with the Laplace transform can be found in Chapter VIII of [146]. ◊

1.2 General Properties of Orthogonal Polynomials

Let ψ be a positive Borel measure on $\mathbb{R}$ with infinite support and let us assume that the moments

$$\mu_n = \int_{\mathbb{R}} x^n d\psi(x), \quad n \geq 0,$$

exist and are finite. We normalize the measure in such a way that $\mu_0 = 1$, so we have a probability measure. Following *Lebesgue's decomposition theorem* any Borel measure on the real line can be decomposed into three measures such that

$$\psi = \psi_c + \psi_d + \psi_{sc},$$

where ψ_c is absolutely continuous, ψ_d is discrete and ψ_{sc} is singular continuous. The absolutely continuous measure ψ_c is classified by the Radon–Nikodym theorem and can always be written (abusing the notation) as $d\psi_c(x) = \psi_c(x)dx,$ with respect to the Lebesgue measure. The discrete measure ψ_d can always be written as

$$d\psi_d(x) = \sum_k \psi(\{x_k\})\delta(x - x_k)\, dx,$$

where k runs over a countable set, x_k are the mass points, $\psi(\{x_k\})$ are the sizes or magnitudes of these jumps and $\delta(x - a)$ is the Dirac delta distribution. Finally, the singular continuous measure ψ_{sc} is defined over a set of measure 0. The *Cantor measure* (the probability measure on the real line whose cumulative distribution function is the Cantor function) is an example of a singular continuous measure. In this text we consider positive Borel measures on $\mathbb{R}$ with either only an absolutely continuous part or only a discrete part (or a combination of both).

Associated with this measure ψ we can consider the Hilbert space L^2_ψ with the inner product

$$(f,g)_\psi = \int_{\mathbb{R}} f(x)g(x)\, d\psi(x), \tag{1.13}$$

of all measurable real functions f such that $(f,f)_\psi = \|f\|^2_\psi < \infty$. If the support of the measure is given by $\mathcal{S} \subseteq \mathbb{R}$, then this space will be written as $L^2_\psi(\mathcal{S})$. When $\mathcal{S}$ is a countable set, for example $\mathbb{N}_0 = \{0, 1, \ldots\}$, this space is usually denoted by $\ell^2_\psi(\mathbb{N}_0)$.

We say that $(p_n(x))_n$ is a *sequence of polynomials* if each element is a polynomial of degree exactly n in the real variable x. A sequence of polynomials is *monic* if the leading coefficient of each polynomial is exactly 1. A sequence of polynomials $(p_n)_n$ is *orthogonal* with respect to a Borel measure ψ if

$$(p_n, p_m)_\psi = \int_{\mathbb{R}} p_n(x)p_m(x)\, d\psi(x) = d_n^2\delta_{nm},$$

where $d_n^2 = \|p_n\|^2_\psi > 0$. If the norm is always identically 1, we say that the polynomial sequence is *orthonormal* and we denote it by $(P_n)_n$. When we work with the sequence of monic orthogonal polynomials, we will use the notation $(\widehat{P}_n)_n$ and its norms will be denoted by $\|\widehat{P}_n\|^2_\psi = \zeta_n$.

Given a Borel measure ψ on $\mathbb{R}$ with infinite support and finite moments, it will always be possible to build a sequence of orthogonal polynomials. A direct way is through the *Gram–Schmidt orthogonalization process* applied to the set $\{1, x, x^2, \ldots\}$.

This method builds the polynomials one by one taking into account that all the previous ones have already been calculated. Specifically

$$\begin{aligned}
\widehat{P}_0(x) &= 1,\\
\widehat{P}_1(x) &= x - \frac{(\widehat{P}_0, x)_\psi}{(\widehat{P}_0, \widehat{P}_0)_\psi}\widehat{P}_0(x),\\
\vdots \qquad & \qquad \vdots\\
\widehat{P}_k(x) &= x^k - \sum_{j=0}^{k-1} \frac{(\widehat{P}_j, x^k)_\psi}{(\widehat{P}_j, \widehat{P}_j)_\psi}\widehat{P}_j(x).
\end{aligned}$$

Once they have been computed, the monic polynomials can be normalized by dividing them by $\|\widehat{P}_k\|_\psi = \sqrt{\zeta_k}$. Observe that the monic orthogonal polynomials have always real coefficients.

Another way to define orthogonal polynomials is through determinants associated with the moments. Consider the determinant

$$\Delta_n = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_n \\ \mu_1 & \mu_2 & \cdots & \mu_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n} \end{vmatrix}, \quad \Delta_{-1} = 1.$$

The quadratic form associated with the matrix of the previous determinant, which we denote by (Δ_n), is always positive definite. Indeed, for any real vector $v = (v_0, v_1, \dots, v_n)^T$, we have that

$$v^T(\Delta_n)v = \sum_{j,k=0}^{n} \mu_{j+k} v_j v_k = \int_{\mathbb{R}} \left[\sum_{j=0}^{n} v_j x^j\right]^2 d\psi(x),$$

which is clearly positive. Thus $\Delta_n > 0, n \geq 0$. $\Delta_n, n \geq 0$ are usually called *Hankel determinants*.

It is easy to see that the sequence of polynomials defined by

$$p_n(x) = \begin{vmatrix} \mu_0 & \mu_1 & \cdots & \mu_{n-1} & 1 \\ \mu_1 & \mu_2 & \cdots & \mu_n & x \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mu_n & \mu_{n+1} & \cdots & \mu_{2n-1} & x^n \end{vmatrix}, \quad n \geq 0, \tag{1.14}$$

is orthogonal with respect to the measure ψ. To see that, we simply evaluate the inner product $(p_n, x^m)_\psi = 0, m = 0, 1, \dots, n-1$ observing that we always have a repeated column, so the determinant is 0. Alternatively, we have $(p_n, x^n)_\psi = \Delta_n > 0$. Thus

$$p_n(x) = \Delta_{n-1} x^n + \text{lower degree terms},$$

and we have that

$$(p_n, p_n)_\psi = (p_n, \Delta_{n-1}x^n)_\psi = \Delta_{n-1}\Delta_n.$$

Therefore, the polynomials

$$P_n(x) = \frac{1}{\sqrt{\Delta_{n-1}\Delta_n}} p_n(x)$$

are orthonormal, and the leading coefficient is given by $h_n = \sqrt{\Delta_{n-1}/\Delta_n} = \zeta_n^{-1/2}$. The monic family can be written as

$$\widehat{P}_n(x) = \frac{1}{\Delta_{n-1}} p_n(x) = \sqrt{\frac{\Delta_n}{\Delta_{n-1}}} P_n(x).$$

Finally, let us see another way to generate the orthogonal polynomials recurrently. Assume that we have a sequence of orthogonal polynomials $(p_n)_n$. The polynomial $xp_n(x)$ has degree $n+1$ and can be expressed as a linear combination of the $n+1$ first polynomials, i.e.

$$xp_n(x) = \sum_{j=0}^{n+1} d_{n,j} p_j(x).$$

Now, multiplying by $p_k(x)$ and evaluating the inner product, it is easy to see, using the orthogonal relations, that the coefficients $d_{n,j} = 0, j = 0, 1, \ldots, n-2$. Therefore, only the last three coefficients remain and every family of orthogonal polynomials satisfies a *three-term recurrence relation* of the form

$$xp_n(x) = a_n p_{n+1}(x) + b_n p_n(x) + c_n p_{n-1}(x), \quad n \geq 0, \quad p_{-1} = 0, \tag{1.15}$$

where

$$a_n = \frac{(xp_n, p_{n+1})_\psi}{(p_{n+1}, p_{n+1})_\psi}, \quad b_n = \frac{(xp_n, p_n)_\psi}{(p_n, p_n)_\psi}, \quad c_n = \frac{(xp_n, p_{n-1})_\psi}{(p_{n-1}, p_{n-1})_\psi}.$$

We observe that the coefficient b_n is always real. Moreover, for the orthonormal family $P_n(x)$ we have, comparing the coefficients of x^{n+1} in (1.15), that $a_n = h_n/h_{n+1} = \sqrt{\zeta_{n+1}/\zeta_n} > 0$, and that $c_n = (xP_n, P_{n-1})_\psi = (P_n, xP_{n-1})_\psi = a_{n-1}$. Therefore the sequence of orthonormal polynomials $(P_n)_n$ satisfies a three-term recurrence relation of the form

$$xP_n(x) = a_n P_{n+1}(x) + b_n P_n(x) + a_{n-1} P_{n-1}(x), \quad a_n > 0, \quad b_n \in \mathbb{R}. \tag{1.16}$$

For the monic family $(\widehat{P}_n)_n$ the three-term recurrence relation will be given by

$$x\widehat{P}_n(x) = \widehat{P}_{n+1}(x) + \alpha_n \widehat{P}_n(x) + \beta_n \widehat{P}_{n-1}(x), \quad \widehat{P}_0(x) = 1, \quad \widehat{P}_1(x) = x - \alpha_0, \tag{1.17}$$

where $\alpha_{n-1} \in \mathbb{R}, \beta_n > 0$ for $n \geq 1$. The relations between these coefficients and the coefficients of the orthonormal family are given by

$$a_n = \sqrt{\frac{\zeta_{n+1}}{\zeta_n}}, \quad \alpha_n = b_n, \quad \beta_n = \frac{\zeta_n}{\zeta_{n-1}}.$$

Observe that $\zeta_n = \beta_n \cdots \beta_1$.

Another way of writing this recurrence relation is in matrix form. Denoting the column vector of orthonormal polynomials by $P(x) = (P_0(x), P_1(x), \ldots)^T$, we have that $xP(x) = JP(x)$, where J is the tridiagonal symmetric matrix

$$J = \begin{pmatrix} b_0 & a_0 & 0 & 0 & 0 & \cdots \\ a_0 & b_1 & a_1 & 0 & 0 & \cdots \\ 0 & a_1 & b_2 & a_2 & 0 & \cdots \\ 0 & 0 & a_2 & b_3 & a_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{1.18}$$

This matrix plays a very important role and it is called a *Jacobi matrix*. In particular, we will find this kind of matrix in the one-step transition probability matrix of a one-dimensional discrete-time birth–death chain and in the infinitesimal operator of a birth–death process, as we will see in the next two chapters. The inverse result, i.e. for a family of polynomials defined by (1.16), where there exists a positive measure for which they are orthogonal, is known as Favard's theorem or the spectral theorem for orthogonal polynomials. We will see more details in Section 1.3.

The powers of J can be computed formally using orthogonality properties. Observe that the relation $xP(x) = JP(x)$ implies that $x^nP(x) = J^nP(x)$. Therefore, multiplying by $P^T(x)$, integrating with respect to the measure ψ and looking at the (i,j) entry, we obtain

$$\int_{\mathbb{R}} x^n P_i(x)P_j(x)d\psi(x) = \sum_{k\geq 0} \int_{\mathbb{R}} J^n_{ik} P_k(x)P_j(x)d\psi(x) = J^n_{ij}. \tag{1.19}$$

From here we observe that the moments $(\mu_n)_n$ of the measure ψ can be computed from J^n_{00}. In general, the diagonal coefficients J^n_{ii} are the moments of the measure $d\psi_i(x) = P_i^2(x)d\psi(x)$.

The identity (1.19) can be extended to any analytic function defined on $\text{supp}(\psi)$ of the form $f(x) = \sum_{n\geq 0} c_n x^n$ as

$$\int_{\mathbb{R}} f(x)P_i(x)P_j(x)\, d\psi(x) = \sum_{n\geq 0} \int_{\mathbb{R}} c_n x^n P_k(x)P_j(x)\, d\psi(x) = \sum_{n\geq 0} c_n J^n_{ij} = f(J)_{ij}.$$