Orthonormal Polynomials, Related Orthonormal Functions and the Hilbert Spaces they Span

Herman J. Bierens

December 5, 2008

1 Orthogonal Polynomials

Let $w(x)$ be a non-negative Borel measurable real-valued function on $\mathbb{R}$ satisfying

$$\int_{-\infty}^{\infty} |x|^k w(x)dx \in (0,\infty) \text{ for } k = 0,1,2,.....$$

where the integral involved is the Lebesgue integral. Without loss of generality we may assume that w is a density function. Let

$$p_k(x|w) = \sum_{j=0}^{k} \alpha_{k,j}x^j,\ \alpha_{k,k} = 1,\ k = 0,1,2,.....,\tag{1}$$

be a sequence of polynomials in $x \in \mathbb{R}$ such that

$$\int_{-\infty}^{\infty} p_k(x|w)p_m(x|w)w(x)dx = 0 \text{ if } k \neq m.\tag{2}$$

In words, the polynomials $p_k(x|w)$ are *orthogonal* with respect to the weight function $w(x)$.

Defining

$$\overline{p}_k(x|w) = \frac{p_k(x|w)}{\sqrt{\int_{-\infty}^{\infty} p_k(y|w)^2 w(y)dy}}\tag{3}$$

yields a sequence of *orthonormal* polynomials w.r.t. $w(x)$:

$$\int_{-\infty}^{\infty} \overline{p}_k(x|w)\overline{p}_m(x|w)w(x)dx = \begin{cases} 0 & \text{if } k \neq m, \\ 1 & \text{if } k = m. \end{cases} \tag{4}$$

This sequence is uniquely determined by $w(x)$, except for signs. In other words, $|\overline{p}_k(x|w)|$ is unique. To show this, suppose that there exists another sequence $\overline{p}_k^*(x|w)$ of orthonormal polynomials w.r.t. $w(x)$. Since $\overline{p}_k^*(x|w)$ is a polynomial of order k, we can write $\overline{p}_k^*(x|w) = \sum_{m=0}^{k} \beta_{m,k}\overline{p}_m(x|w)$. Similarly, we can write $\overline{p}_k(x|w) = \sum_{m=0}^{k} \alpha_{m,.k}\overline{p}_m^*(x|w)$. Then for $j < k$,

$$\begin{aligned} \int_{-\infty}^{\infty} \overline{p}_k^*(x|w)\overline{p}_j(x|w)w(x)dx &= \sum_{m=0}^{j} \alpha_{m,j} \int_{-\infty}^{\infty} \overline{p}_k^*(x|w)\overline{p}_m^*(x|w)w(x)dx \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} \int_{-\infty}^{\infty} \overline{p}_k^*(x|w)\overline{p}_j(x|w)w(x)dx &= \sum_{m=0}^{k} \beta_{m,k} \int_{-\infty}^{\infty} \overline{p}_m(x|w)\overline{p}_j(x|w)w(x)dx \\ &= \beta_{j,k} \int_{-\infty}^{\infty} \overline{p}_j(x|w)^2 w(x)dx = \beta_{j,k}, \end{aligned}$$

hence $\beta_{j,.k} = 0$ for $j < k$ and thus

$$\overline{p}_k^*(x|w) = \beta_{k,k}\overline{p}_k(x|w).$$

Moreover, by normality,

$$1 = \int_{-\infty}^{\infty} \overline{p}_k^*(x|w)^2 w(x)dx = \beta_{k,k}^2 \int_{-\infty}^{\infty} \overline{p}_k(x|w)^2 w(x)dx = \beta_{k,k}^2,$$

so that $\overline{p}_k^*(x|w) = \pm\overline{p}_k(x|w)$. Consequently, $|\overline{p}_k(x|w)|$ is unique.

2 The Hilbert Space Spanned by Orthonormal Polynomials

The reason for considering orthonormal polynomials is the following.

Theorem 1. *Let* $w(x)$ *be a density function with support* (a,b), $-\infty \leq a < b \leq \infty$, *satisfying the moment conditions*

$$\int_a^b |x|^k \, w(x)dx < \infty \tag{5}$$

for $k = 0, 1, 2, ...$ *Let* $L_w^2\,(a,b)$ *be the Hilbert space of Borel measurable real functions* f *on* (a,b) *satisfying* $\int_a^b f(x)^2 w(x)dx < \infty$, *with inner product* $\langle f, g\rangle = \int_a^b f(x)g(x)w(x)dx$ *and associated norm* $\|f\| = \sqrt{\langle f, f\rangle}$ *and metric* $||f-g||$. *For an arbitrary function* $f \in L_w^2\,(a,b)$, *let*

$$f_n(x) = \sum_{k=0}^{n} \gamma_k \overline{p}_k(x|w), \ n \in \mathbb{N},$$

where for $k \in \mathbb{N}$,[1]

$$\gamma_k = \langle f, \overline{p}_k\rangle = \int_a^b f(x)\overline{p}_k(x|w)w(x)dx \tag{6}$$

Then

$$\lim_{n\to\infty} \|f - f_n\| = 0. \tag{7}$$

This result implies that every function $f \in L_w^2\,(a,b)$ *can be written as*

$$f(x) = \sum_{k=0}^{\infty} \gamma_k \overline{p}_k(x|w) \ \textit{a.e. on} \ (a,b). \tag{8}$$

Proof: Appendix.

Note that condition (5) holds trivially if the support of $w(x)$ is bounded: $\max(|a|, |b|) < \infty$. However, as is well-known, condition (5) also holds for the standard normal density, the exponential density and more generally the density of the Gamma distribution, for example.

Since for every density $w(x)$ with support (a,b), $\int_a^b f(x)^2 \, dx < \infty$ implies that $f(x)/\sqrt{w(x)} \in L_w^2\,(a,b)$, the following corollary of Theorem 1 holds trivially.

---

[1] $\mathbb{N} = \{0, 1, 2, 3, ...\}$.

Corollary 1. *Let $L^2(a,b)$ be the Hilbert space of square integrable Borel measurable real functions on (a,b), with inner product $\langle f,g\rangle = \int_a^b f(x)\,g(x)\,dx$ and associated norm and metric. Every function $f \in L^2(a,b)$ can be written as*

$$f(x) = \sqrt{w(x)}\left(\sum_{k=0}^{\infty}\gamma_k \overline{p}_k(x)\right) \text{ a.e. on } (a,b),$$

where $w \in \mathcal{M}(a,b)$, the $\overline{p}_k(x)$'s are the orthonormal polynomials generated by $w(x)$ and the γ_k's are the Fourier coefficients of $f(x)/\sqrt{w(x)}$, i.e.,

$$\gamma_k = \int_a^b f(x)\overline{p}_k(x)\sqrt{w(x)}dx.$$

This result implies that the functions

$$\psi_k(x|w) = \overline{p}_k(x|w)\sqrt{w(x)},\ w \in \mathcal{M}(a,b),\ k \in \mathbb{N},$$

form a complete *orthonormal sequence*[2] *in $L^2(a,b)$.*

Of course, the $\psi_k(x|w)$'s are no longer polynomials.

If $\max(|a|,|b|) < \infty$ then there is another way to construct a complete orthonormal sequence in $L^2(a,b)$, as follows. Let $W(x)$ be the distribution function of a density w with bounded support (a,b). Then

$$G(x) = a + (b-a)W(x)$$

is a one-to-one mapping of (a,b) onto (a,b), with inverse

$$G^{-1}(y) = W^{-1}((y-a)/(b-a))$$

where W^{-1} is the inverse of $W(x)$. For every $f \in L^2(a,b)$,

$$(b-a)\int_a^b f(G(x))^2 w(x)dx = \int_a^b f(G(x))^2 dG(x) = \int_a^b f(x)^2 dx < \infty.$$

Hence $f(G(x)) \in L^2(w)$ and thus by Theorem 1,

$$f(G(x)) = \sum_{k=0}^{\infty}\gamma_k\overline{p}_k(x) \text{ a.e. on } (a,b)$$

---

[2] See for example Young (1988) for the definition of complete orthonormal sequences.

where

$$\begin{aligned}\gamma_k &= \int_a^b f\left(G\left(x\right)\right)\overline{p}_k(x|w)w(x)dx = \frac{1}{b-a}\int_a^b f\left(G\left(x\right)\right)\overline{p}_k(x|w)dG(x) \\ &= \frac{1}{b-a}\int_a^b f\left(x\right)\overline{p}_k\left(G^{-1}\left(x\right)|w\right)dx\end{aligned}$$

Consequently

$$f\left(x\right) = f\left(G\left(G^{-1}\left(x\right)\right)\right) = \sum_{k=0}^{\infty}\gamma_k\overline{p}_k\left(G^{-1}\left(x\right)\right) \text{ a.e. on } (a,b)$$

Note that $dG^{-1}\left(x\right)/dx = dG^{-1}\left(x\right)/dG\left(G^{-1}\left(x\right)\right) = 1/G'\left(G^{-1}\left(x\right)\right)$, so that

$$\begin{aligned}&\int_a^b \overline{p}_k\left(G^{-1}\left(x\right)|w\right)\overline{p}_m\left(G^{-1}\left(x\right)|w\right)dx \\ &= \int_a^b \overline{p}_k\left(G^{-1}\left(x\right)|w\right)\overline{p}_m\left(G^{-1}\left(x\right)|w\right)G'\left(G^{-1}\left(x\right)\right)dG^{-1}\left(x\right) \\ &= \int_a^b \overline{p}_k\left(x|w\right)\overline{p}_m\left(x|w\right)G'\left(x\right)dx \\ &= (b-a)\int_a^b \overline{p}_k\left(x|w\right)\overline{p}_m\left(x|w\right)w\left(x\right)dx = (b-a)\,I\left(k=m\right)\end{aligned}$$

Thus,

Corollary 2. *Let* w be a density with bounded support (a,b). *Let* W *be the c.d.f. of a density* w*, with inverse* W^{-1}*. Then the functions*

$$\psi_k\left(x|w\right) = \overline{p}_k\left(W^{-1}\left(\left(x-a\right)/\left(b-a\right)\right)|w\right)/\sqrt{(b-a)},\ k\in\mathbb{N},$$

form a complete orthonormal sequence in $L^2\left(a,b\right)$*, i.e., every* $f\in L^2\left(a,b\right)$ *can be written as* $f\left(x\right) = \sum_{k=0}^{\infty}\alpha_k\psi_k\left(x|w\right)$ *a.e. on* (a,b)*, where* $\alpha_k = \int_a^b f\left(x\right)$ $\psi_k\left(x|w\right)dx$*.*

3 The Three-Term Recurrence Relation

It follows from (1) that $p_0(x|w)\equiv 1$, and it follows from (2) that $p_1(x|w) = \alpha_{1,0}+x$ can be constructed by solving $\int_{-\infty}^{\infty}\left(\alpha_{1,0}+x\right)w(x)dx = \alpha_{1,0}\int_{-\infty}^{\infty}w(x)$

$dx + \int_{-\infty}^{\infty} x.w(x)dx = 0$, hence

$$\alpha_{1,0} = -\int_{-\infty}^{\infty} x.w(x)dx \Big/ \int_{-\infty}^{\infty} w(x)dx.$$

The question now arises how to construct these orthogonal polynomials further for $k \geq 2$.

The answer is the following.

Theorem 2. *Every sequence of polynomials* $p_k(x|w) = \sum_{j=0}^{k} \alpha_{k,j}x^j$, *with* $\alpha_{k,k} = 1$, *satisfying the orthogonality condition* (2), *can be generated recursively by the three-term recurrence relation* (*hereafter referred to as TTRR*)

$$p_{k+1}(x|w) + (b_k - x)\,p_k(x|w) + c_k p_{k-1}(x|w) = 0,\ \ k \geq 1, \tag{9}$$

where

$$b_k = \frac{\int_{-\infty}^{\infty} x.p_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx} \tag{10}$$

and

$$c_k = \frac{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_{k-1}(x|w)^2 w(x)dx} \tag{11}$$

Proof: Appendix.

Note that b_k is invariant for the normalization $\alpha_{k,k} = 1$, but c_k is not. In the case

$$\underline{p}_k(x|w) = \sum_{j=0}^{k} \alpha_{k,j}x^j,\ \ \alpha_{k,k} \neq 0,$$

$$\int \underline{p}_k(x|w)\underline{p}_m(x|w)w(x)dx = 0 \text{ for } k \neq m$$

we have $p_k(x|w) = \underline{p}_k(x|w)/\alpha_{k,k}$, so that

$$\underline{c}_k = \frac{\int_{-\infty}^{\infty} \underline{p}_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} \underline{p}_{k-1}(x|w)^2 w(x)dx} = \frac{\alpha_{k,k}^2}{\alpha_{k-1,k-1}^2} c_k$$

The TTRR for $\underline{p}_k(x|w)$ now becomes

$$a_k \underline{p}_{k+1}(x|w) + (b_k - x)\,\underline{p}_k(x|w) + \underline{c}_k.a_{k-1}\underline{p}_{k-1}(x|w) = 0,\ \ k \geq 1,$$

where $a_k = \alpha_{k1,k}/\alpha_{k+1,k+1} = \lim_{x\to\infty} x.\underline{p}_k(x|w)/\underline{p}_{k+1}(x|w)$. Thus,

Theorem 3. *Without the normalization $\alpha_{k,k} = 1$ in (1) the TTRR for $p_k(x|w)$ becomes*

$$a_k p_{k+1}(x|w) + (b_k - x)\, p_k(x|w) + c_k p_{k-1}(x|w) = 0, \;\; k \geq 1, \tag{12}$$

where

$$a_k = \lim_{|x|\to\infty} \frac{x.p_k(x|w)}{p_{k+1}(x|w)}, \;\; b_k = \frac{\int_{-\infty}^{\infty} x.p_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}$$

and

$$c_k = a_{k-1} \frac{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_{k-1}(x|w)^2 w(x)dx}.$$

It follows now from Theorem 3 that:

Theorem 4. *Every sequence $\overline{p}_k(x|w)$ of orthonormal polynomials with respect to a weight function $w(x)$ can be generated recursively by the TTRR*

$$a_{k+1}.\overline{p}_{k+1}(x|w) + (b_k - x)\, \overline{p}_k(x|w) + a_k.\overline{p}_{k-1}(x|w) = 0, \quad k \geq 1. \tag{13}$$

where

$$a_k = \left| \lim_{|x|\to\infty} \frac{x.\overline{p}_{k-1}(x|w)}{\overline{p}_k(x|w)} \right| \tag{14}$$

and

$$b_k = \int_{-\infty}^{\infty} x.\overline{p}_k(x|w)^2 w(x)dx. \tag{15}$$

Proof: Appendix.

4 Examples of Orthonormal Polynomials

4.1 Hermite Polynomials

If $w(x)$ is the density of the standard normal distribution, the orthonormal polynomials involved satisfy the TTRR

$$\sqrt{k+1}\overline{p}_{k+1}(x|w) - \frac{1}{\sqrt{k+1}} x.\overline{p}_k(x|w) + \sqrt{k}\overline{p}_{k-1}(x|w) = 0, \quad k \geq 1,$$

starting from $\overline{p}_0(x|w) = 1$, $\overline{p}_1(x|w) = x$. Thus in this case $a_k = \sqrt{k}$ and $b_k = 0$ in (13). These polynomials are known as Hermite[3] polynomials.

It follows from Theorem 1 that the Hermite polynomials span the Hilbert space $L^2_w(\mathbb{R})$, where $w(x) = \exp\left(-x^2/2\right)/\sqrt{2\pi}$. Consequently, any density $f(x)$ on $\mathbb{R}$ can be represented by

$$f(x) = w(x)\left(\sum_{k=0}^{\infty} \gamma_k \overline{p}_k(x|w)\right)^2$$

where $\sum_{k=0}^{\infty} \gamma_k^2 = 1$.

Gallant and Nychka (1987) propose to use these polynomials to generalize the standard normal density to a semi-nonparametric (SNP) density function

$$f_n(x) = w(x)\left(\sum_{k=0}^{n} \gamma_{k,n} \overline{p}_k(x|w)\right)^2,$$

where $\sum_{k=0}^{n} \gamma_{k,n}^2 = 1$.

4.2 Laguerre Polynomials

The standard exponential density function

$$w(x) = I\left(x \geq 0\right)\exp\left(-x\right),$$

where $I(.)$ is the indicator function, gives rise to the orthonormal Laguerre[4] polynomials, with TTRR

$$(k+1)\,\overline{p}_{k+1}(x|w) + (2k+1-x)\,\overline{p}_k(x|w) + k.\overline{p}_{k-1}(x|w) = 0, \quad k \geq 1,$$

starting from $\overline{p}_0(x|w) = 1$, $\overline{p}_1(x|w) = x - 1$. Thus in this case $a_k = k$ and $b_k = 2k + 1$.

4.3 Legendre Polynomials

The uniform density on $[-1, 1]$,

$$w(x) = \frac{1}{2} I\left(|x| \leq 1\right),$$

---

[3] Charles Hermite (1822-1901).

[4] Edmund Nicolas Laguerre (1834-1886)

generates the orthonormal Legendre[5] polynomials on $(-1,1)$, with TTRR

$$\begin{aligned}\frac{k+1}{\sqrt{2k+3}\sqrt{2k+1}}\overline{p}_{k+1}(x|w) - x.\overline{p}_k(x|w) \\ + \frac{k}{\sqrt{2k+1}\sqrt{2k-1}}\overline{p}_{k-1}(x|w) = 0, k \geq 1,\end{aligned} \tag{16}$$

starting from $\overline{p}_0(x|w) = 1$, $\overline{p}_1(x|w) = \sqrt{3}x$.

4.4 Shifted Legendre polynomials

Note that the orthonormal Legendre polynomials $\overline{p}_k(x|w)$ satisfy

$$\begin{aligned}&\int_0^1 \overline{p}_k(2u-1|w)\overline{p}_m(2u-1|w)du \\ &\quad = \frac{1}{2}\int_0^1 \overline{p}_k(2u-1|w)\overline{p}_m(2u-1|w)d\left(2u-1\right) \\ &\quad = \frac{1}{2}\int_{-1}^1 \overline{p}_k(x|w)\overline{p}_m(x|w)dx = I\left(k=m\right)\end{aligned}$$

Hence,

$$\rho_k(u) = \overline{p}_k(2u-1|w),\ k = 0,1,2,....,$$

is a sequence of orthonormal polynomials w.r.t. the uniform density on $[0,1]$,

$$w(u) = I\left(0 \leq u \leq 1\right)$$

The $\rho_k(u)$'s are known as the shifted Legendre polynomials, also called the orthonormal Legendre polynomials on the unit interval $[0,1]$. Substituting $x = 2u-1$ and $\overline{p}_k(x) = \rho_k(u)$ in (16) yields the TTRR

$$\begin{aligned}\frac{\left(k+1\right)/2}{\sqrt{2k+3}\sqrt{2k+1}}\rho_{k+1}(u) + \left(0.5-u\right).\rho_k(u) \\ + \frac{k/2}{\sqrt{2k+1}\sqrt{2k-1}}\rho_{k-1}(u) = 0,\ k \geq 1,\end{aligned}$$

starting from $\rho_0(u) = 1$, $\rho_1(u) = \sqrt{3}\left(2u-1\right)$.

---

[5] Adrien-Marie Legendre (1752-1833)

These Legendre polynomials have been used by Bierens (2007) and Bierens and Carvalho (2007) to model semi-nonparametrically the unobserved heterogeneity of interval-censored mixed proportional hazard models and bivariate mixed proportional hazard models, respectively.

In particular, the conditional survival function of a mixed proportional hazard model, proposed by Lancaster(1979), takes the form

$$S(t|X) = P[T > t|X] = E\left[\exp\left(-V.\varphi\left(X\right)\Lambda(t)\right)\middle| X\right]$$

where T is a duration, X is a vector of covariates, $\varphi\left(X\right) > 0$ is the systematic hazard possibly depending on parameters,[6] $\Lambda(t) = \int_0^t \lambda(\tau)d\tau$ is the integrated baseline hazard, with $\lambda(t) \geq 0$ the baseline hazard, and $V > 0$ represents the unobserved heterogeneity, which is assumed to be independent of X.

With $G(v)$ the unknown distribution function of V, $S(t|X)$ takes the form

$$S(t|X) = \int_0^\infty \exp\left(-v.\varphi\left(X\right)\Lambda(t)\right)dG(v) = H\left(\exp\left(-\varphi\left(X\right)\Lambda(t)\right)\right)$$

where

$$H(u) = \int_0^\infty u^v dG(v),\ u \in [0,1],$$

is a distribution function on $[0,1]$, with density

$$h(u) = \int_0^\infty v u^{v-1} dG(v). \tag{17}$$

The corresponding conditional density function of T is

$$\begin{aligned}\psi(t|X) &= -\partial S(t|X)/\partial t \\ &= h\left(\exp\left(-\varphi\left(X\right)\Lambda(t)\right)\right)\exp\left(-\varphi\left(X\right)\Lambda(t)\right)\varphi\left(X\right)\lambda(t)\end{aligned}$$

Elbers and Ridder (1982) have shown that under the normalization

$$E[V] = \int_0^\infty v dG(v) = 1 \tag{18}$$

and some mild regularity conditions the MPH model is nonparametrically identified. Note that the condition (18) is equivalent to $h(1) = 1$.

---

[6] Usually $\varphi\left(X\right)$ is parametrized as $\varphi\left(X\right) = \exp(\beta' X)$.

Bierens (2007) has proposed to model the density $h(u)$ semi-nonparametrically, using orthonormal Legendre polynomials $\rho_k(u)$ on $[0,1]$, as

$$h_n(u) = \frac{\left(1+\sum_{k=1}^{n}\delta_k\rho_k(u)\right)^2}{1+\sum_{k=1}^{n}\delta_k^2}. \tag{19}$$

This is motivated by the fact that for any density function $h(u)$ on $[0,1]$ there exists a sequence δ_k satisfying $\sum_{k=1}^{\infty}\delta_k^2 < \infty$ such that

$$h(u) = \frac{\left(1+\sum_{k=1}^{\infty}\delta_k\rho_k(u)\right)^2}{1+\sum_{k=1}^{\infty}\delta_k^2} \tag{20}$$

4.5 Chebyshev Polynomials

Chebyshev polynomials are generated by the weight function

$$w(x) = \frac{1}{\pi\sqrt{1-x^2}} I\left(|x| < 1\right).$$

This is a density function on $(-1,1)$. To see this, let $\theta = \arccos(x)$, so that $x = \cos(\theta)$, and observe that

$$\frac{dx}{d\theta} = -\sin(\theta) = -\sqrt{1-\cos^2(\theta)} = -\sqrt{1-x^2},$$

hence

$$\frac{d\arccos(x)}{dx} = \frac{-1}{\sqrt{1-x^2}} \tag{21}$$

Then

$$\begin{aligned}\int_{-1}^{1}\frac{1}{\pi\sqrt{1-x^2}}dx &= -\frac{1}{\pi}\int_{-1}^{1} d\arccos(x) \\ &= \frac{\arccos(-1)-\arccos(1)}{\pi} = 1\end{aligned}$$

because $\arccos(-1) = \pi$ and $\arccos(1) = 0$.

The orthogonal (but not orthonormal) Chebyshev polynomials $p_k(x|w)$ satisfy the TTRR

$$p_{k+1}(x|w) - 2xp_k(x|w) + p_{k-1}(x|w) = 0,\ k \geq 1, \tag{22}$$

starting from $p_0(x|w) = 1$, $p_1(x|w) = x$, with orthogonality properties

$$\int_{-1}^{1} \frac{p_k(x|w)p_m(x|w)}{\pi\sqrt{1-x^2}}dx = \begin{cases} 0 & \text{if} \quad k \neq m, \\ 1/2 & \text{if} \quad k = m > 0, \\ 1 & \text{if} \quad k = m = 0. \end{cases}$$

An important practical difference with the other polynomials discussed so far is that Chebyshev polynomials have a closed form[7]:

$$p_k(x|w) = \cos\left(k.\arccos(x)\right). \tag{23}$$

To see this, observe from (21) and the well-known sine-cosine formulas that

$$\begin{aligned}
&\int_{-1}^{1} \frac{\cos\left(k.\arccos(x)\right)\cos\left(m.\arccos(x)\right)}{\pi\sqrt{1-x^2}}dx \\
&= -\frac{1}{\pi}\int_{-1}^{1} \cos\left(k.\arccos(x)\right)\cos\left(m.\arccos(x)\right) d\arccos(x) \\
&= \frac{1}{\pi}\int_{0}^{\pi} \cos\left(k.\theta\right)\cos\left(m.\theta\right) d\theta \\
&= \frac{1}{2\pi}\int_{0}^{\pi} \cos\left(\left(k+m\right)\theta\right) d\theta + \frac{1}{2\pi}\int_{0}^{\pi} \cos\left(\left(k-m\right)\theta\right) d\theta \\
&= \frac{1}{2}\left(\frac{\sin\left(\left(k+m\right)\pi\right)}{\left(k+m\right)\pi} + \frac{\sin\left(\left(k-m\right)\pi\right)}{\left(k-m\right)\pi}\right) \\
&= \begin{cases} 0 & \text{if} \quad k \neq m, \\ 1/2 & \text{if} \quad k = m > 0, \\ 1 & \text{if} \quad k = m = 0. \end{cases}
\end{aligned} \tag{24}$$

Moreover, the TTRR (22) follows from

$$\begin{aligned}
&\cos\left(\left(k+1\right).\theta\right) - 2\cos(\theta)\cos\left(k.\theta\right) + \cos\left(\left(k-1\right).\theta\right) \\
&= \cos\left(k.\theta\right)\cos(\theta) - \sin\left(k.\theta\right)\sin(\theta) - 2\cos(\theta)\cos\left(k.\theta\right) \\
&+ \cos\left(k.\theta\right)\cos(\theta) + \sin\left(k.\theta\right)\sin(\theta) = 0.
\end{aligned}$$

Hence, the functions (23) satisfy the TTRR (22) and are therefore genuine polynomials.

---

[7] Note that $\arccos(x) = \operatorname{atan}\left(-x/\sqrt{1-x^2}\right) + \frac{1}{2}.\pi$, where $\operatorname{atan}(x)$ is the inverse of the tangents function $\tan(\theta) = \sin(\theta)/\cos(\theta)$, $\theta \in (-\pi/2, \pi/2)$. In most programming languages the function $\operatorname{atan}(x)$ is an intrinsic function. For example, in Visual Basic this function is the $\mathrm{ATN}(x)$ function.

In view of (24) we can now define the orthonormal Chebyshev polynomials as

$$C_k(x) = \begin{cases} 1 & \text{for} \quad k = 0, \\ \sqrt{2}\cos\left(k.\arccos(x)\right) & \text{for} \quad k \geq 1. \end{cases}$$

Thus, it follows from Corollary 1 that every function $f \in L^2\left(-1,1\right)$ can be written as

$$f\left(x\right) = \frac{\sum_{k=0}^{\infty} \gamma_k C_k(x)}{\pi\sqrt{1-x^2}} \text{ a.e. on } (-1,1),$$

where

$$\gamma_k = \int_{-1}^{1} \frac{f\left(x\right) C_k(x)}{\sqrt{\pi\sqrt{1-x^2}}} dx$$

4.6 Shifted Chebyshev polynomials

Similar to the shifted Legendre polynomials, let for $u \in [0,1]$,

$$\overline{C}_k(u) = \begin{cases} 1 & \text{for} \quad k = 0 \\ \sqrt{2}.\cos\left(k.\arccos(2u-1)\right) & \text{for} \quad k \geq 1 \end{cases} \tag{25}$$

and let

$$w_c(u) = \frac{2}{\pi\sqrt{1-(2u-1)^2}} = \frac{1}{\pi\sqrt{u\left(1-u\right)}}. \tag{26}$$

Observe from (21) that

$$w_c(u) = \frac{1}{\pi}\frac{d\left(-\arccos(2u-1)\right)}{du} \tag{27}$$

so that the c.d.f. of $w_c(u)$ is

$$W_c(u) = \int_0^u w_c(x)dx = 1 - \pi^{-1}\arccos(2u-1). \tag{28}$$

Moreover, it is easy to verify from (24) that

$$\int_0^1 \overline{C}_k(u)\overline{C}_m(u)w_c(u)du = \begin{cases} 0 & \text{if} \quad k \neq m, \\ 1 & \text{if} \quad k = m. \end{cases}$$

5 Cosine Series Approximation of Functions in $L^2(0,1)$

5.1 Chebyshev Approximation of Functions in $L^2(0,1)$

Consider the Hilbert space $L^2_{w_c}(0,1)$ spanned by the shifted Chebyshev polynomials (25). Recall from Theorem 1 that $L^2_{w_c}(0,1)$ contains all Borel measurable real functions $f(u)$ on $[0,1]$ satisfying

$$\int_0^1 f(u)^2 w_c(u)du < \infty \tag{29}$$

where $w_c(u)$ is defined by (26).

The condition (29) is stronger a condition than in the case of the Hilbert space $L^2(0,1)$ because $\min_{0\leq u\leq 1} w_c(u) = w_c(1/2) = 2/\pi$, so that (29) implies

$$\int_0^1 f(u)^2 du \leq \frac{\pi}{2}\int_0^1 f(u)^2 w_c(u)du < \infty,$$

but not the other way around. Thus, $L^2_{w_c}(0,1) \subset L^2(0,1)$.

On the other hand, it follows from Corollary 1 that every function $f \in L^2(0,1)$ can be written as

$$\begin{aligned} f(u) &= \sqrt{w_c(u)}\sum_{k=0}^{\infty}\gamma_k \overline{C}_k(u) \\ &= \sqrt{w_c(u)}\left(\gamma_0 + \sum_{k=1}^{\infty}\gamma_k\sqrt{2}\cos\left(k.\arccos(2u-1)\right)\right) \end{aligned} \tag{30}$$

a.e. on $(0,1)$, where

$$\gamma_k = \int_0^1 \left(f(u)/\sqrt{w_c(u)}\right)\overline{C}_k(u)w_c(u)du = \int_0^1 f(u)\overline{C}_k(u)\sqrt{w_c(u)}du.$$

5.2 Cosine Series Representation

It is easy to verity that the c.d.f. $W_c(u)$ defined by (28) has inverse

$$W_c^{-1}(u) = \frac{1}{2}\left(1+\cos\left(\pi\left(1-u\right)\right)\right)$$

It follows therefore from Corollary 2 that every function $f \in L^2(0,1)$ can be written as

$$\begin{aligned} f(u) &= \gamma_0 + \sum_{k=1}^{\infty} \gamma_k \sqrt{2} \cos\left(k.\arccos(2W_c^{-1}(u)-1)\right) \\ &= \gamma_0 + \sum_{k=1}^{\infty} \gamma_k \sqrt{2} \cos\left(k\pi(1-u)\right) \\ &= \gamma_0 + \sum_{k=1}^{\infty} \gamma_k (-1)^k \sqrt{2} \cos(k\pi u) \\ &= \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \sqrt{2} \cos(k\pi u) \end{aligned}$$

where

$$\begin{aligned} \alpha_k &= \gamma_k (-1)^k = (-1)^k \int_0^1 f(u) \overline{C}_k \left(\frac{1}{2}(1+\cos(\pi(1-u))) \right) du \\ &= \begin{cases} \int_0^1 f(u)\, du & \text{if} \quad k=0, \\ \int_0^1 f(u) \sqrt{2} \cos(k\pi u)\, du & \text{if} \quad k \geq 1. \end{cases} \end{aligned}$$

In other words,

Theorem 5. *The functions*

$$\kappa_k(u) = \begin{cases} 1 & if \quad k=0, \\ \sqrt{2}\cos(k\pi u) & if \quad k \geq 1, \end{cases}$$

form a complete orthonormal sequence in $L^2(0,1)$*. Thus, given a function* $f \in L^2(0,1)$*, let*

$$f_n(u) = \alpha_0 + \sum_{k=1}^{n} \alpha_k \sqrt{2} \cos(k\pi u)$$

be the projection of f *on* $\kappa_0, \kappa_0, ..., \kappa_0$*. Then* $\alpha_k = \int_0^1 f(u) \kappa_k(u)\, du$, $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$*, and*

$$\lim_{n\to\infty} \int_0^1 (f(u) - f_n(u))^2\, du = \lim_{n\to\infty} \sum_{k=n+1}^{\infty} \alpha_k^2 = 0.$$

Consequently,[8] *f can be written as*

$$f(u) = \alpha_0 + \sum_{k=1}^{\infty} \alpha_k \sqrt{2} \cos(k\pi u) \ \text{ a.e. on } (0,1).$$

5.3 Comparison with Classical Fourier Analysis

As is well-known[9], the functions

$$1, \ \sqrt{2}\cos(k\pi x), \ \sqrt{2}\sin(k\pi x), \ k = 1, 2, 3, ...$$

form a complete orthonormal sequence in $L^2(-1,1)$ with respect to the weight function $w(x) = \frac{1}{2} I(|x| \leq 1)$. Thus, for every real function $g \in L^2(-1,1)$,

$$\lim_{n\to\infty} \int_{-1}^{1} (g(x) - g_n(x))^2 \, dx = 0. \tag{31}$$

where

$$g_n(x) = \alpha_0 + \sum_{k=1}^{n} \alpha_k \sqrt{2}\cos(k\pi x) + \sum_{k=1}^{n} \beta_k \sqrt{2}\sin(k\pi x) \tag{32}$$

with

$$\begin{aligned}
\alpha_0 &= \frac{1}{2}\int_{-1}^{1} g(x)\, dx \\
\alpha_k &= \frac{1}{2}\int_{-1}^{1} \sqrt{2}\cos(k\pi x)\, g(x)\, dx \\
\beta_k &= \frac{1}{2}\int_{-1}^{1} \sqrt{2}\sin(k\pi x)\, g(x)\, dx \\
\sum_{k=1}^{\infty} \alpha_k^2 &< \infty, \ \sum_{k=1}^{\infty} \beta_k^2 < \infty.
\end{aligned}$$

Let $f(u) \in L^2(0,1)$ be arbitrary, and let $g(x) = f(|x|)$. Then $g(x) \in L^2(-1,1)$, with Fourier coefficients

$$\alpha_0 = \frac{1}{2}\int_{-1}^{1} f(|x|)\, dx = \int_{0}^{1} f(u)\, du$$

---

[8] Similar to Theorem 1.

[9] See for example Young (1988)

$$\begin{aligned}\alpha_k &= \frac{1}{2}\int_{-1}^{1}\sqrt{2}\cos\left(k\pi x\right)f\left(|x|\right)dx = \int_{0}^{1}\sqrt{2}\cos\left(k\pi u\right)f\left(u\right)du \\ \beta_k &= \frac{1}{2}\int_{-1}^{1}\sqrt{2}\sin\left(k\pi x\right)f\left(|x|\right)dx = 0\end{aligned}$$

Hence it follows from (31) that

$$\begin{aligned}&\lim_{n\to\infty}\int_{0}^{1}\left(f\left(u\right)-\alpha_0-\sum_{k=1}^{n}\alpha_k\sqrt{2}\cos\left(k\pi u\right)\right)^2 du \qquad (33)\\ &= \frac{1}{2}\lim_{n\to\infty}\int_{-1}^{1}\left(f\left(|x|\right)-\alpha_0-\sum_{k=1}^{n}\alpha_k\sqrt{2}\cos\left(k\pi x\right)\right)^2 dx \\ &= 0\end{aligned}$$

Similar to the proof of Theorem 1 is follows now from (33) that

$$f\left(u\right) = \alpha_0 + \sum_{k=1}^{\infty}\alpha_k\sqrt{2}\cos\left(k\pi u\right) \text{ a.e. on } (0,1),$$

where $\alpha_0 = \int_0^1 f\left(u\right)du$ and $\alpha_k = \int_0^1 \sqrt{2}\cos\left(k\pi u\right)f\left(u\right)du$ for $k \geq 1$, which is just the result in Theorem 5.

6 Trigonometric Representation of Density and Distribution Functions

6.1 Density Functions on the Unit Interval

It follows from Theorem 5 that for any density function $h(u)$ on $[0,1]$ there exists a sequence $\{\alpha_k\}_{k=0}^{\infty}$ satisfying $\sum_{k=0}^{\infty}\alpha_k^2 = 1$ such that

$$h\left(u\right) = \left(\alpha_0 + \sum_{k=1}^{\infty}\alpha_k\sqrt{2}\cos\left(k\pi u\right)\right)^2 \text{ a.e. on } (0,1). \qquad (34)$$

However, similar to Bierens (2007) it follows that the α_k's in (34) are no longer unique. For example, we can always write $h\left(u\right) = f_p\left(u\right)^2$ where for arbitrary $p \in [0,1]$,

$$f_p\left(u\right) = \left(I\left(u < p\right) - I\left(u \geq p\right)\right)\sqrt{h(u)}. \qquad (35)$$

Then the α_k's in (34) take the form

$$\alpha_k = \int_0^p \sqrt{h(u)}\kappa_k(u)\,du - \int_p^1 \sqrt{h(u)}\kappa_k(u)\,du$$

In particular, we may choose for α_0 any

$$\alpha_0 \in \left[-\int_0^1 \sqrt{h(u)}du, \int_0^1 \sqrt{h(u)}du\right]. \tag{36}$$

If we choose $\alpha_0 \in \left(0, \int_0^1 \sqrt{h(u)}du\right]$ then we can reparametrize the Fourier coefficients α_k as

$$\begin{aligned} \alpha_0 &= \frac{1}{\sqrt{1+\sum_{m=1}^{\infty}\delta_m^2}} \\ \alpha_k &= \frac{\delta_k}{\sqrt{1+\sum_{m=1}^{\infty}\delta_m^2}},\ k \geq 1 \end{aligned}$$

where $\sum_{m=0}^{\infty}\delta_m^2 < \infty$. Hence,

Theorem 6. *For any density function $h(u)$ on $[0,1]$ there exists a sequence $\{\delta_m\}_{m=1}^{\infty}$ satisfying $\sum_{m=0}^{\infty}\delta_m^2 < \infty$ such that*

$$h(u) = \frac{\left(1+\sum_{k=1}^{\infty}\delta_k\sqrt{2}\cos(k\pi u)\right)^2}{1+\sum_{m=1}^{\infty}\delta_m^2} \text{ a.e. on } (0,1). \tag{37}$$

Moreover, the corresponding SNP densities

$$h_n(u) = \frac{\left(1+\sqrt{2}\sum_{k=1}^{n}\delta_k\cos(k\pi u)\right)^2}{1+\sum_{m=1}^{n}\delta_m^2} \tag{38}$$

satisfy

$$\lim_{n\to\infty}\int_0^1 |h(u)-h_n(u)|\,du = 0. \tag{39}$$

Furthermore, the corresponding SNP distribution function $H_n(u) = \int_0^u h_n(v)dv$ has the closed form expression

$$H_n(u) = u + \frac{1}{1+\sum_{m=1}^{n}\delta_m^2}$$

$$\times\left[2\sqrt{2}\sum_{k=1}^{n}\delta_k\frac{\sin\left(k\pi u\right)}{k\pi}+\sum_{k=1}^{n}\delta_k^2\frac{\sin\left(2k\pi u\right)}{2k\pi}\right.$$

$$\left.+\sum_{m=1}^{n-1}\sum_{k=m+1}^{n}\delta_m\delta_k\left(\frac{\sin\left(\left(k+m\right)\pi u\right)}{\left(k+m\right)\pi}+\frac{\sin\left(\left(k-m\right)\pi u\right)}{\left(k-m\right)\pi}\right)\right] \tag{40}$$

The latter result follows from the well-known sine-cosine formulas, and (39) follows from the corresponding result in Bierens (2007) based on shifted Legendre polynomials.

6.2 General Representation of Density and Distribution Functions

Given a continuous distribution function $G(x)$ with support $\Xi \subset \mathbb{R}$, any distribution function $F(x)$ with support contained in Ξ can be written as $F(x) = H(G(x))$, where $H(u) = F(G^{-1}(u))$ is a distribution function on $[0,1]$. Moreover, if F and G are absolutely continuous with densities f and g, respectively, then H is absolutely continuous with density $h(u)$, and $f(x) = h(G(x))g(x)$. Therefore, $f(x)$ can be estimated semi-nonparametrically by estimating $h(u)$ semi-nonparametrically. The role of G is twofold. First, G determines the support of f. Second, G acts as an initial guess of the unknown distribution function F. Obviously, in the latter case the initial guess is right if H is the c.d.f. of the uniform distribution on $[0,1]$.

Therefore, it follows from Theorem 6 that

Theorem 7. *Given an absolutely continuous distribution function $G(x)$ on $\mathbb{R}$ with density $g(x)$, any density function $f(x)$ with support contained in the support of g* (i.e., $\{x : f(x) > 0\} \subset \{x : g(x) > 0\}$) *can be written as*

$$f(x) = g(x)\frac{\left(1+\sqrt{2}\sum_{k=1}^{\infty}\delta_k\cos\left(k\pi G\left(x\right)\right)\right)^2}{1+\sum_{m=1}^{\infty}\delta_m^2} \tag{41}$$

a.e. on $\{x : f(x) > 0\}$. Moreover, the corresponding SNP densities

$$f_n(x) = g(x)\frac{\left(1+\sqrt{2}\sum_{k=1}^{n}\delta_k\cos\left(k\pi G\left(x\right)\right)\right)^2}{1+\sum_{m=1}^{n}\delta_m^2} \tag{42}$$

satisfy

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}|f(x)-f_n(x)|\,dx=0.$$

7 References

Bierens, H.J. (2004), *Introduction to the Mathematical and Statistical Foundations of Econometrics.* Cambridge University Press.

Bierens, H.J. (2007), "Semi-Nonparametric Interval-Censored Mixed Proportional Hazard Models: Identification and Consistency Results", forthcoming in *Econometric Theory.*

Bierens, H.J. & J. R. Carvalho (2007), "Semi-Nonparametric Competing Risks Analysis of Recidivism", *Journal of Applied Econometrics* 22, 971-993.

Elbers, C. & G. Ridder (1982), "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model", *Review of Economic Studies* 49, 403-409.

Gallant, A. R. & D. W. Nychka (1987), "Semi-Nonparametric Maximum Likelihood Estimation", *Econometrica* 55, 363-390.

Lancaster, T. (1979), "Econometric Methods for the Duration of Unemployment", *Econometrica* 47, 939-956.

Young, N. (1988), *An Introduction to Hilbert Space.* Cambridge University Press.

8 Appendix

8.1 Proof of Theorem 1

Let $\overline{f}_n(x)=\sum_{k=0}^{n}\gamma_k\overline{p}_k(x|w)$, where $\gamma_k=\int_a^b\overline{p}_k(x|w)f(x)w(x)dx$, and observe that due to condition (5), $\overline{f}_n\in L_w^2(a,b)$. Next, observe that

$$\begin{aligned}\left\|f-\overline{f}_n\right\|^2 &= \int_a^b\left(f(x)-\sum_{k=0}^{n}\gamma_k\overline{p}_k(x|w)\right)^2 w(x)dx\\ &= \int_a^b f(x)^2w(x)dx-2\sum_{k=0}^{n}\gamma_k\int_a^b\overline{p}_k(x|w)f(x)w(x)dx\end{aligned}$$

$$
\begin{aligned}
&+\sum_{k_1=0}^{n}\sum_{k_1=0}^{n}\gamma_{k_1}\gamma_{k_2}\int_a^b \overline{p}_{k_1}(x|w)\overline{p}_{k_2}(x|w)w(x)dx \\
&= \int_a^b f(x)^2 w(x)dx - \sum_{k=0}^{n}\gamma_k^2 \geq 0. \qquad (43)
\end{aligned}
$$

Hence $\sum_{k=0}^{n}\gamma_k^2 \leq \int_a^b f(x)^2 w(x)dx < \infty$ for all $n \geq 0$, and thus

$$\sum_{k=0}^{\infty}\gamma_k^2 < \infty. \qquad (44)$$

The latter implies that $\overline{f}_n$ is a Cauchy sequence in $L_w^2(a,b)$ because

$$\lim_{\min(n,m)\to\infty}\left\|\overline{f}_n - \overline{f}_m\right\|^2 = \lim_{\min(n,m)\to\infty}\sum_{k=\min(n,m)+1}^{\max(n,m)}\gamma_k^2 \leq \lim_{m\to\infty}\sum_{k=m}^{\infty}\gamma_k^2 = 0.$$

It follows therefore from the definition of a Hilbert space[10] that there exists a function $\overline{f} \in L_w^2(a,b)$ such that

$$\lim_{n\to\infty}\left\|\overline{f}_n - \overline{f}\right\| = 0. \qquad (45)$$

This limit function $\overline{f}$ can be written as

$$\overline{f}(x) = \sum_{k=0}^{n}\gamma_k\overline{p}_k(x|w) + \varepsilon_n(x) \qquad (46)$$

for all $n \in \mathbb{N}$, where

$$\lim_{n\to\infty}\int_a^b \varepsilon_n(x)^2 w(x)dx = 0. \qquad (47)$$

Proof of (7): To prove (7), it suffices to show that

$$\int_a^b \exp(i.t.x)\left(f(x) - \overline{f}(x)\right)w(x)dx = 0 \qquad (48)$$

for all $t \in R$, because (48) implies that $f(x) = \overline{f}(x)$ a.e. on (a,b), due to the uniqueness of the Fourier transform.[11] .

---

[10] In particular the property that every Cauchy sequence in a Hilbert space converges to a limit contained in this Hilbert space.

[11] See for example Bierens (1994, Theorem 3.1.1, p.50).

It follows from the definition of γ_m and $\overline{f}$ that for $m \leq n$,

$$\begin{aligned}\left|\int_a^b \left(f(x)-\overline{f}\,(x)\right)\overline{p}_m(x|w)w(x)dx\right| &= \left|\int_a^b \varepsilon_n(x)\overline{p}_m(x|w)w(x)dx\right| \\ &\leq \sqrt{\int_a^b \varepsilon_n(x)^2 w(x)dx},\end{aligned}$$

hence by (47),

$$\int_a^b \left(f(x)-\overline{f}\,(x)\right)\overline{p}_m(x|w)w(x)dx = 0 \tag{49}$$

for all $m \in \mathbb{N}$. This result implies, by induction, that

$$\int_a^b \left(f(x)-\overline{f}\,(x)\right)x^m w(x)dx = 0 \text{ for all } m \in \mathbb{N}. \tag{50}$$

In its turn (50) implies, together with the well-known equality $\exp(i.t.x) = \sum_{m=0}^{\infty}(i.t.x)^m/m!$, that for $t \in \mathbb{R}$ and all $n \in \mathbb{N}$,

$$\begin{aligned}&\int_a^b \exp(i.t.x)\left(f(x)-\overline{f}\,(x)\right)w(x)dx \\ &\quad= \int_a^b \sum_{m=0}^{n}\frac{(i.t.x)^m}{m!}\left(f(x)-\overline{f}\,(x)\right)w(x)dx \\ &\quad+ \int_a^b \left(\sum_{m=n+1}^{\infty}\frac{(i.t.x)^m}{m!}\right)\left(f(x)-\overline{f}\,(x)\right)w(x)dx \\ &\quad= \int_a^b \left(\sum_{m=n+1}^{\infty}\frac{(i.t.x)^m}{m!}\right)\left(f(x)-\overline{f}\,(x)\right)w(x)dx\end{aligned}$$

If $-\infty < a < b < \infty$ then by dominated convergence,

$$\begin{aligned}&\int_a^b \exp(i.t.x)\left(f(x)-\overline{f}\,(x)\right)w(x)dx \\ &\quad= \int_a^b \left(\lim_{n\to\infty}\sum_{m=n+1}^{\infty}\frac{(i.t.x)^m}{m!}\right)\left(f(x)-\overline{f}\,(x)\right)w(x)dx = 0\end{aligned}$$

If $a=-\infty$ and/or $b=\infty$ we can find for arbitrary $\varepsilon>0$ a finite lower bound $a(\varepsilon)>a$ and a finite upper bound $b(\varepsilon)<b$ such that

$$\begin{aligned}\left|\int_a^{a(\varepsilon)}\exp(i.t.x)\left(f(x)-\overline{f}\,(x)\right)w(x)dx\right| &< \varepsilon/2\\ \left|\int_{b(\varepsilon)}^{b}\exp(i.t.x)\left(f(x)-\overline{f}\,(x)\right)w(x)dx\right| &< \varepsilon/2\end{aligned}$$

whereas by dominated convergence

$$\begin{aligned}&\int_{a(\varepsilon)}^{b(\varepsilon)}\exp(i.t.x)\left(f(x)-\overline{f}\,(x)\right)w(x)dx\\ &\quad=\int_{a(\varepsilon)}^{b(\varepsilon)}\left(\lim_{n\to\infty}\sum_{m=n+1}^{\infty}\frac{(i.t.x)^m}{m!}\right)\left(f(x)-\overline{f}\,(x)\right)w(x)dx=0\end{aligned}$$

Since $\varepsilon>0$ is arbitrary, we therefore have in either case that (48) holds. It therefore follows from (46) and (47) that

$$\lim_{n\to\infty}\int_a^b\left(f\left(x\right)-\sum_{k=0}^{n}\gamma_k\overline{p}_k(x|w)\right)^2 w\left(x\right)dx=0. \tag{51}$$

This completes the proof of (7)

Proof of (8): To prove that (7) implies (8), let X be a random drawing from $w(x)$. Then by Chebyshev's inequality, (51) implies

$$f\left(X\right)=p\lim_{n\to\infty}\sum_{k=0}^{n}\gamma_k\overline{p}_k(X|w) \tag{52}$$

As is well-known[12], convergence in probability is equivalent to almost sure (a.s.) convergence along a further subsequence of an arbitrary subsequence of n. Thus it follows from (52) that for any subsequence n_j in $\mathbb{N}$ there exists a further subsequence n_{j_m} such that for $m\to\infty$,

$$\sum_{k=0}^{n_{j_m}}\gamma_k\overline{p}_k(X|w)\to f(X)\text{ a.s.} \tag{53}$$

---

[12] See for example Bierens (2004, Theorem 6.B.3, p.168).

For each n there exists an m such that $n_{j_{m-1}} \leq n < n_{j_m}$. Hence, there exists a further subsequence j_n of n_{j_m} such that for $n \to \infty$,

$$\sum_{k=0}^{j_n} \gamma_k \overline{p}_k(X|w) \to f(X) \text{ a.s.,} \tag{54}$$

and $j_{n-1} \leq n < j_n$. The latter implies that

$$\begin{aligned} E\left[\left(\sum_{k=0}^{j_n} \gamma_k \overline{p}_k(X|w) - \sum_{k=0}^{n} \gamma_k \overline{p}_k(X|w)\right)^2\right] &= E\left(\sum_{k=n+1}^{j_n} \gamma_k \overline{p}_k(X|w)\right)^2 \\ &\leq \sum_{k=j_{n-1}+1}^{j_n} \gamma_k^2 \end{aligned}$$

so that

$$\begin{aligned} \sum_{n=1}^{\infty} E\left[\left(\sum_{k=0}^{j_n} \gamma_k \overline{p}_k(X|w) - \sum_{k=0}^{n} \gamma_k \overline{p}_k(X|w)\right)^2\right] &\leq \sum_{n=1}^{\infty} \sum_{k=j_{n-1}+1}^{j_n} \gamma_k^2 \\ &\leq \sum_{k=0}^{\infty} \gamma_k^2 < \infty \end{aligned}$$

Then by Chebyshev's inequality,

$$\sum_{n=1}^{\infty} P\left[\left|\sum_{k=0}^{j_n} \gamma_k \overline{p}_k(X|w) - \sum_{k=0}^{n} \gamma_k \overline{p}_k(X|w)\right| > \varepsilon\right] < \infty$$

for all $\varepsilon > 0$, which by the Borel-Cantelli lemma[13] implies that for $n \to \infty$

$$\sum_{k=0}^{j_n} \gamma_k \overline{p}_k(X|w) - \sum_{k=0}^{n} \gamma_k \overline{p}_k(X|w) \to 0 \text{ a.s.} \tag{55}$$

Combining (54) and (55), it follows that $\sum_{k=0}^{n} \gamma_k \overline{p}_k(X|w) \to f(X)$ a.s. as $n \to \infty$, which is equivalent to (8) because the support of $w(x)$ was assumed to be (a, b). Q.E.D.

---

[13] See for example Bierens (2004, Theorem 2.B.2, p. 168).

8.2 Proof of Theorem 2

Due to the normalization $\alpha_{k,k} = 1$ it follows that $p_{k+1}(x|w) - x.p_k(x|w)$ is a polynomial of order k, which can be written as a linear combination of $p_0(x|w), p_1(x|w),...,p_k(x|w)$:

$$p_{k+1}(x|w) - x.p_k(x|w) = \sum_{j=0}^{k} \delta_{j,k} p_j(x|w) \tag{56}$$

for example. Then for $m \leq k$,

$$\begin{aligned} 0 &= \int_{-\infty}^{\infty} p_{k+1}(x|w)p_m(x|w)w(x)dx - \int_{-\infty}^{\infty} x.p_k(x|w)p_m(x|w)w(x)dx \\ &\quad - \sum_{j=0}^{k} \delta_{j,k} \int_{-\infty}^{\infty} p_j(x|w)p_m(x|w)w(x)dx \\ &= -\int_{-\infty}^{\infty} x.p_k(x|w)p_m(x|w)w(x)dx - \delta_{m,k} \int_{-\infty}^{\infty} p_m(x|w)^2 w(x)dx \end{aligned}$$

so that

$$\delta_{m,k} = -\frac{\int_{-\infty}^{\infty} x.p_k(x|w)p_m(x|w)w(x)dx}{\int_{-\infty}^{\infty} p_m(x|w)^2 w(x)dx}$$

Because $x.p_m(x|w)$ is a polynomial of order $m+1$, it follows that for $m \leq k-2$, $x.p_m(x|w)$ is orthogonal to $p_k(x|w)$, hence $\delta_{m,k} = 0$ for $m = 0, 1, ..., k-2$. Thus it follows from (56) that

$$\begin{aligned} p_{k+1}(x|w) - x.p_k(x|w) &= \delta_{k,k} p_k(x|w) + \delta_{k-1,k} p_{k-1}(x|w) \\ &= -b_k p_k(x|w) - c_k p_{k-1}(x|w) \end{aligned}$$

where

$$b_k = -\delta_{k,k} = \frac{\int_{-\infty}^{\infty} x.p_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}$$

and

$$\begin{aligned} c_k &= -\delta_{k-1,k} = \frac{\int_{-\infty}^{\infty} x.p_k(x|w)p_{k-1}(x|w)w(x)dx}{\int_{-\infty}^{\infty} p_{k-1}(x|w)^2 w(x)dx} \\ &= \frac{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_{k-1}(x|w)^2 w(x)dx} \end{aligned}$$

The last equality is easy to verify.

8.3 Proof of Theorem 4

Denote $d_k = \sqrt{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}$ and $p_k(x|w) = d_k.\overline{p}_k(x|w)$. It follows trivially from (12) that

$$a_k \frac{d_{k+1}}{d_k} \overline{p}_{k+1}(x|w) + (b_k - x)\,\overline{p}_k(x|w) + c_k \frac{d_{k-1}}{d_k} \overline{p}_{k-1}(x|w) = 0,\ k \geq 1. \tag{57}$$

However, with a_k and c_k defined in Theorem 3 we have

$$d_k/d_{k-1} = \sqrt{\frac{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_{k-1}(x|w)^2 w(x)dx}} = \sqrt{\frac{c_k}{a_{k-1}}},$$

so that (57) can be written as

$$\sqrt{c_{k+1}.a_k.}\overline{p}_{k+1}(x|w) + (b_k - x)\,\overline{p}_k(x|w) + \sqrt{c_k.a_{k-1}}\overline{p}_{k-1}(x|w) = 0,\ k \geq 1.$$

Note that

$$\begin{aligned} \sqrt{c_{k+1}.a_k} &= \left| \lim_{|x|\to\infty} \frac{x.p_k(x|w)}{p_{k+1}(x|w)} \right| \sqrt{\frac{\int_{-\infty}^{\infty} p_{k+1}(x|w)^2 w(x)dx}{\int_{-\infty}^{\infty} p_k(x|w)^2 w(x)dx}} \\ &= \left| \lim_{|x|\to\infty} \frac{x.\overline{p}_k(x|w)}{\overline{p}_{k+1}(x|w)} \right|. \end{aligned}$$

Redefining $\sqrt{c_{k+1}.a_k}$ as the new a_{k+1} in (14), the TTRR (13) follows.