On the rate of convergence of the laws of Markov chains associated with orthogonal polynomials

Marc Lindlbauer *

Institute for Biomathematics and Biometry, GSF – National Research Center for Environment and Health, Ingolstädter Landstraße 1, D-85764 Neuherberg, Germany

Received 1 November 1997; received in revised form 20 March 1998

Abstract

We investigate random walks \((S_n)_{n \in \mathbb{N}_0}\) on the nonnegative integers arising from isotropic random walks on distance transitive graphs. The laws of those isotropic random walks converge in distribution to the normal distribution and the transition probabilities of the \(S_n\) are closely related with a sequence of Bernstein–Szegő polynomials. We give an explicit representation for these polynomials as a sum of Chebychev polynomials of the second kind and using this representation we prove an upper bound for the rate of convergence of the laws of the \(S_n\). © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Random walks; Orthogonal polynomials; Limit theorems

1. Introduction

The central limit theorem for sums of \(\mathbb{R}\)-valued independent and identically distributed random variables and the theorem of Berry–Esséen are well known from probability theory. In 1990 Voit [9] proved analogous results in a very general setting for random walks \((S_n)_{n \in \mathbb{N}_0}\) on \(\mathbb{N}_0\) whose transition probabilities are closely connected with certain sequences of orthogonal polynomials. Using the methods from [9] we investigate the rate of convergence in the central limit theorem for the \(S_n\) for the special case where these random variables arise from isotropic random walks on an infinite distance transitive graph which is a generalization of homogeneous trees. Closely related with the structure of infinite distance transitive graphs is a sequence \((P_n)_{n \in \mathbb{N}_0}\) of orthogonal polynomials which turn out to be Bernstein–Szegő polynomials. We give an explicit representation of the \(P_n\) as a sum

* E-mail: lindlb@gsf.de.
of Chebychev polynomials of the second kind that enables us to improve the rate of convergence presented in [9]. Furthermore we compare the rate in this paper with the classical rate from the theorem of Berry–Esseen. The paper is organized as follows: Since this paper is based on the setting in [9] we give a short overview of it in Section 2. Section 3 deals with random walks on infinite distance transitive graphs which are the motivating example for how the theorems are stated. Section 4 introduces some notation which is necessary to state the main results in Section 5. Section 6 presents the proofs for the results in Section 5.

2. Random walks associated with a sequence of orthogonal polynomials

Let \((a_n)_{n\in\mathbb{N}}, (b_n)_{n\in\mathbb{N}}\) and \((c_n)_{n\in\mathbb{N}}\) be sequences of real numbers satisfying \(a_n, c_n > 0, b_n \geq 0\) and \(a_n + b_n + c_n = 1\). Furthermore, we assume that \(\alpha := \lim_{n \to \infty} a_n, \beta := \lim_{n \to \infty} b_n\) and \(\gamma := \lim_{n \to \infty} c_n\) exist and that

(i) \(0 < \gamma < \alpha < 1\) and
(ii) \(\sum_{n=1}^{\infty} n \cdot \max(0, b_n - b_{n+1}) < \infty\).

Condition (i) is essential for this paper since for \(\alpha = \gamma\) the asymptotic behaviour of the laws of the \(S_n\) is of a very different kind (see, e.g., [4, 10]). Condition (ii) is a technical one that is necessary to obtain the results in [9]. Both conditions are stated for completeness and will implicitly be used in this paper when applying [9].

Using Favard’s theorem (see, e.g., [2]) we can define a sequence of orthogonal polynomials by

\[
P_0 \equiv 1, \quad P_1(x) = 2\sqrt{\alpha} \cdot x + \beta, \quad P_1 \cdot P_n = a_n P_{n+1} + b_n P_n + c_n P_{n-1}.
\]

Since the \((P_n)_{n\in\mathbb{N}_0}\) form a basis of the complex vector space of polynomials in one complex variable, the linearization coefficients of the product \(P_m \cdot P_n = \sum_{k=|m-n|}^{m+n} g_{m,n,k} P_k\) are uniquely determined and we can define a convolution of point measures \(\delta_m\) and \(\delta_n\) on \(\mathbb{N}_0\) for all \(m, n \in \mathbb{N}_0\) by setting

\[
\delta_m \ast \delta_n := \sum_{k=|m-n|}^{m+n} g_{m,n,k} \delta_k.
\]

We point out that the range of the summation index in (2) is due to the orthogonality of the \(P_n\). Moreover, if we assume that all the linearization coefficients \(g_{m,n,k}\) are nonnegative then we can extend the convolution \(\ast\) of point measures uniquely to a norm continuous convolution on \(M_b(\mathbb{N}_0)\), the space of all bounded measures on \(\mathbb{N}_0\). With this convolution \(\mathbb{N}_0\) can be given the structure of a polynomial hypergroup. For details see [6, 7, 11]. If in the following a convolution on \(M_b(\mathbb{N}_0)\) occurs it will always be assumed to arise in the described way and will be denoted as \(\ast_p\) to specify the defining sequence of orthogonal polynomials.

In the following let \((S_n)_{n\in\mathbb{N}_0}\) be a \(\mathbb{N}_0\)-valued Markov chain defined on a probability space \((\Omega, \mathcal{A}, \mathbb{P})\). Using the above convolution we introduce:
Definition 2.1. A Markov chain \((S_n)_{n \in \mathbb{N}_0}\) is called homogeneous with respect to the convolution \(*_p\) if there exists a sequence \((\nu_n)_{n \in \mathbb{N}}\) of probability measures on \(\mathbb{N}_0\) with

\[
\mathbb{P}(S_{n+1} = r \mid S_n = k) = \nu_{n+1} *_p \delta_r(\{r\})
\]

for all \(n, r, k \in \mathbb{N}_0\). The \(\nu_n\) are called transition measures. If all \(\nu_n\) are equal to a probability measure \(\nu\) then \((S_n)_{n \in \mathbb{N}_0}\) is called stationary.

Since the results of this paper do not depend on the law of \(S_0\) (see [9]) we may assume without loss of generality that \(S_0 \equiv 0\). An easy induction on \(n\) yields the following important consequence of the definition:

Proposition 2.2. For all \(n \in \mathbb{N}\) the law of \(S_n\) is given by \(\nu_1 *_p \nu_2 *_p \cdots *_p \nu_n\).

Finally we remark that the normalization \(P_0 = 1\) is a necessary condition for a family of orthogonal polynomials to define a polynomial hypergroup. The choice of \(P_1\) is due to Voit (see, e.g., [9] or [10]).

3. Distance transitive graphs and the associated random walks on \(\mathbb{N}_0\)

Let \(\Gamma\) denote an undirected, connected graph with countably infinite vertex set \(V\). Furthermore we assume that \(\Gamma\) is locally finite, i.e. that every vertex has only a finite number of neighbours. Let \(d\) denote the usual graph-theoretic distance. With this notation we can define as follows:

Definition 3.1. Let \(\Gamma\) be as above with automorphism group \(A\). \(\Gamma\) is called distance transitive if for all \(u, v, w, y \in V\) with \(d(u, v) = d(w, y)\) there exists \(g \in A\) such that \(g(u) = w\) and \(g(v) = y\).

MacPherson (see [8]) showed that such graphs must have a very simple geometry:

Theorem 3.2 (MacPherson). Every infinite locally finite distance transitive graph is a standard graph \(\Gamma_{a, b}\) with \(a, b \in \mathbb{N} \setminus \{1\}\).

A standard graph \(\Gamma_{a, b}\) is a graph where exactly \(a\) copies of a complete graph with \(b\) vertices are tacked to every vertex of \(\Gamma\). An example for \(a = 2\) and \(b = 3\) is given in Fig. 1.

Now consider a probability space \((\Omega, \mathcal{A}, \mathbb{P})\) and a simple random walk \((T_n)_{n \in \mathbb{N}_0}\) on \(\Gamma\) starting at a fixed vertex \(v_0 \in V\), which is also a Markov chain. The property to be a simple random walk means the following:

1. \(\mathbb{P}(T_0 = v_0) = 1\),
2. \(\mathbb{P}(T_{n+1} = u \mid T_n = v) = \mathbb{P}(T_{n+1} = g(u) \mid T_n = g(v))\), \((u, v \in V, n \in \mathbb{N}_0, g \in A)\),
3. \(\mathbb{P}(d(T_{n+1}, T_n) > 1) = 0\) for all \(n \in \mathbb{N}_0\).

Because of (2) such a random walk is called isotropic. By setting \(S_n := d(T_n, v_0)\) we obtain a \(\mathbb{N}_0\)-valued random walk, that is again a Markov chain. Since properties (2) and (3) imply that in one step the random walk \(T_n\) can only move to its neighbours with equal probability it can easily
be seen using the above characterization of MacPherson of infinite distance transitive graphs that the transition probabilities of the \( S_n \) are given as follows:

\[
\mathbb{P}(S_{n+1} = r \mid S_n = k) = \begin{cases} 
\frac{a-1}{a}, & r = k + 1, \; k \neq 0, \\
\frac{b-2}{a(b-1)}, & r = k, \; k \neq 0, \\
\frac{1}{a(b-1)}, & r = k - 1, \; k \neq 0, \\
1, & r = 1, \; k = 0, \\
0, & |r - k| > 1,
\end{cases}
\]

for all \( n,k,r \in \mathbb{N}_0 \). We note that property (2) ensures that the above transition probabilities do not depend on \( v_0 \). Using these probabilities we define a sequence \( (C_n)_{n \in \mathbb{N}_0} \) of orthogonal polynomials as follows:

\[
C_0 \equiv 1, \quad C_1(x) = 2\sqrt{\frac{a-1}{a(b-1)}} \cdot x + \frac{b-2}{a(b-1)}, \\
C_1 \cdot C_n = \frac{a-1}{a} C_{n+1} + \frac{b-2}{a(b-1)} C_n + \frac{1}{a(b-1)} C_{n-1}.
\]

Computing the orthogonality measure of these polynomials it turns out that they are Bernstein–Szegö polynomials. It is given by the following formula (see, e.g., [1, (4.28)–(4.30)]):

\[
\int_{-1}^{1} C_m(x)C_n(x)w(x) \, dx = 0 \quad (m \neq n, \; a \geq b \geq 2), \\
\int_{-1}^{1} C_m(x)C_n(x)w(x) \, dx + l(a,b) \cdot C_m(x_1)C_n(x_1) = 0 \quad (m \neq n, \; b \geq a \geq 2),
\]
where
\[ w(x) = \frac{\sqrt{1 - x^2}}{2\pi (x_1 - x)(x_2 - x)}, \quad l(a, b) = \frac{b - a}{ab}, \]
\[ x_1 = \frac{2 - a - b}{2\sqrt{(a - 1)(b - 1)}}, \quad x_2 = \frac{ab - a - b + 2}{2\sqrt{(a - 1)(b - 1)}}. \]

Following [9] we will call the \( C_n \) generalized Cartier polynomials. Since all the linearization coefficients in the product \( C_m C_n = \sum_{k=|m-n|}^{m+n} g_{m,n,k} C_k \) are nonnegative (see [9]) the Cartier polynomials define a convolution structure on \( M_b(\mathbb{N}_0) \) as described in Section 2. Therefore the transition probabilities of a random walk arising from a simple isotropic random walk on infinite distance transitive graphs are given for all \( n, k, r \in \mathbb{N}_0 \) by \( P(S_{n+1} = r | S_n = k) = \delta_{1+c} \delta_k(\{r\}) \) for all \( n, k, r \in \mathbb{N}_0 \). Hence \( (S_n)_{n \in \mathbb{N}_0} \) is a stationary random walk on \( \mathbb{N}_0 \) according to Definition 2.1. We generalize this situation replacing the transition measure \( \delta_1 \) by an arbitrary probability measure \( \nu \). In the context of isotropic random walks on \( \Gamma \) this amounts to dropping condition (3). Finally we note that except for the trivial case \( a = b = 2 \) all assumptions of Section 2 are satisfied.

4. Technical notes

In order to state the main results of this paper we have to introduce some notation. If \( f \) is an arbitrary \( \mathbb{C} \)-valued function on \( \mathbb{N}_0 \) and if \( \nu \) is an arbitrary probability measure let \( \nu(f) \) denote the integral \( \int_{\mathbb{N}_0} f(k)\nu(dk) \) whenever this integral exists. Using the notation of Section 2 we define:
\[ x_0 := \frac{1 - \beta}{2\sqrt{\frac{2}{\gamma}}} > 1, \quad \theta_0 := \ln(\sqrt{\frac{2}{\gamma}}) > 0. \]

\( x_0 \) and \( \theta_0 \) are related by the formula \( x_0 = \cos(i\theta_0) \). From (1) in Section 2 we obtain by an easy induction argument that \( P_n(x_0) = 1 \) for all \( n \in \mathbb{N}_0 \). Next we present a sequence of functions which will be essential for our proof for the rate of convergence.

**Definition 4.1.** For all \( n \in \mathbb{N}_0 \) we define a function \( m_n \) on \( \mathbb{N}_0 \) by
\[ m_n(k) := i^n \left( \frac{d}{d\theta} \right)^n P_k(\cos(\theta)) \bigg|_{\theta = i\theta_0}. \]

Those functions are called moment functions of \( n \)th order.

The following proposition lists some important properties of the moment functions \( m_1 \) and \( m_2 \). The proof is given in [9]:
Proposition 4.2. (i) $m_1$ and $m_2$ are non-negative.
(ii) $(m_1(k))^2 < m_2(k)$ for all $k \in \mathbb{N}$ and $m_1(0) = m_2(0) = 0$.
(iii) If $v_1$ and $v_2$ are probability measures then
\[
\begin{align*}
v_1 *_{p_r} v_2(m_1) &= v_1(m_1) + v_2(m_1), \\
v_1 *_{p_r} v_2(m_2) &= v_1(m_2) + v_2(m_2) + 2v_1(m_1)v_2(m_1).
\end{align*}
\]

For the generalized Cartier polynomials the moment function $m_1$ can be computed explicitly (see [9]). For all $n \in \mathbb{N}_0$ we have
\[
m_1(n) = n - \frac{b(a-1)}{a(ab-a-b)} \left(1 - \frac{1}{(a-1)^r(b-1)^r}\right).
\]

If $Y$ is any $\mathbb{N}_0$-valued random variable and $\mathbb{E}(m_1(Y)) < \infty$ and $\mathbb{E}(m_2(Y)) < \infty$ we can define
\[
\mathbb{E}_{m}(Y) := \mathbb{E}(m_1(Y)), \quad \mathbb{V}_{m}(Y) := \mathbb{E}(m_2(Y)) - \mathbb{E}_{m}(Y)^2
\]
to be the modified expectation and the modified variance, respectively. As a consequence of Proposition 4.2(ii) we have $\mathbb{V}_{m}(Y) > 0$ and $\mathbb{V}_{m}(Y) > 0$ if the law of $Y$ is not equal to $\delta_0$. Now let $(S_n)_{n \in \mathbb{N}_0}$ be a stationary Markov chain with respect to $*_{r_p}$ with transition measure $v$. If we assume $v(m_1)$ and $v(m_2)$ to exist we have $\mathbb{E}_{m}(S_n) = n \mathbb{E}_{m}(S_1)$ and $\mathbb{V}_{m}(S_n) = n \mathbb{V}_{m}(S_1)$. This is an immediate consequence of Proposition 4.2(iii) and Proposition 2.2.

5. The main results

Let $(S_n)_{n \in \mathbb{N}_0}$ again be a stationary Markov chain with transition measure $v \neq \delta_0$. Furthermore we assume that $\mathbb{V}(m_1) < \infty$ and $\mathbb{V}(m_2) < \infty$.

Analogous to sums of independent identically distributed $\mathbb{R}$-valued random variables we define a random variable $S_n$ for all $n \in \mathbb{N}_0$ by $S_n := \frac{m_1(S_n) - n \mathbb{E}_{m}(S_1)}{\sqrt{n \mathbb{V}_{m}(S_1)}} S_n$ as in the previous section. Since $\mathbb{V}_{m}(S_1) > 0$ due to $v \neq \delta_0$ division by $\sqrt{\mathbb{V}_{m}(S_1)}$ is allowed. Voit [9] showed that the sequence of the $S_n$ converges in distribution to the standardized normal distribution $\mathcal{N}(0, 1)$ and that an analogue of the theorem of Berry–Esséen holds. The theorem below improves the results of [9] for the special case of generalized Cartier polynomials. The next proposition assures that the assumptions of Theorem 5.2 make sense, i.e. that there are examples where the assumptions are satisfied.

Proposition 5.1. For the Cartier polynomials we have the following representation as a sum of Chebychev polynomials of the second kind for all $n \in \mathbb{N}$ and all $\theta \in \mathbb{C}\setminus\{0\}$:
\[
C_n(\cos \theta) = (a - 1) \frac{\sin((n + 1)\theta) + \frac{b}{(a-1)(b-1)^{\frac{n-2}{2}}} \sin(n\theta) - \frac{1}{a-1} \sin((n - 1)\theta)}{\frac{1}{a[(a-1)(b-1)]^{\frac{n-2}{2}}} \sin(\theta)}.
\]
As a consequence we have
\[ \sup_{n \in \mathbb{N}_0} |C_n(\cos(i\theta_0 - x)) - e^{i\delta_1(n)x}| = O(x^2) \]
for \( x \in \mathbb{R} \) with \( |x| \to 0 \).

Essential for the proof of the theorem below is to have a bound for the above supremum. So it is natural to ask how fast the \( S_n \) converge to \( \mathcal{N}(0, 1) \) in the setting of Section 2 provided that we know how \( \sup_{n \in \mathbb{N}_0} |P_n(\cos(i\theta_0 - x)) - e^{i\delta_1(n)x}| \) behaves. An answer is given in the following theorem:

**Theorem 5.2.** Let \((P_n)_{n \in \mathbb{N}_0}\) be a sequence of orthogonal polynomials as in Section 2. Furthermore let \((S_n)_{n \in \mathbb{N}_0}\) denote a stationary Markov chain on \( \mathbb{N}_0 \) with transition measure \( v \) satisfying \( v \neq \delta_0 \) and \( \sum_{k \in \mathbb{N}_0} k^3 v(k) < \infty \). Set \( \mu := E_{P_n}(S_1) \) and \( \sigma^2 := V_{P_n}(S_1) \) and let \( F_n \) and \( G \) be the distribution functions of \( (m_1(S_n) - n\mu)/\sqrt{n \sigma^2} \) and \( \mathcal{N}(0, 1) \), respectively. If we assume that \( \sup_{n \in \mathbb{N}_0} |P_n(\cos(i\theta_0 - x)) - e^{i\delta_1(n)x}| = O(x^r) \) \((r \geq 0)\) the following holds:

\[ \|F_n - G\|_\infty = O(n^{r/2} \bar{n}) \]

If the convolution \(*_p\) arises from Cartier polynomials we have \( \|F_n - G\|_\infty = O(n^{-1/2}) \).

**Remark.** The crucial point in the proof of Theorem 5.2 is the existence of an upper bound for \( \sup_{n \in \mathbb{N}_0} |P_n(\cos(i\theta_0 - x)) - e^{i\delta_1(n)x}| \). The only way known to the author to obtain such a bound is using Taylor expansion of \( P_n(\cos(i\theta_0 - x)) - e^{i\delta_1(n)x} \) as is done in the proof of Proposition 5.1. In this case we can almost reach the classical rate \( O(n^{-1/2}) \) provided we have a polynomial bound of sufficient high order.

6. Proof of the main results

**Proof of Proposition 5.1.** From the recurrence formula for the generalized Cartier polynomials we have

\[ C_1(x)C_n(x) = \frac{2}{a} \sqrt{\frac{a-1}{b-1}} xC_n(x) + \frac{b-2}{a(b-1)} C_n(x) \]

and hence

\[ \frac{2}{a} \sqrt{\frac{a-1}{b-1}} xC_n(x) = \frac{a-1}{a} C_{n+1}(x) + \frac{1}{a(b-1)} C_{n-1}. \] (5)

With

\[ r := \frac{1}{a} \sqrt{\frac{a-1}{b-1}} \quad k := \frac{b-2}{a(b-1)}, \quad a_n := \frac{a-1}{a} \quad \text{and} \quad c_n := \frac{1}{a(b-1)} \]
one can define a sequence \( Q_n \) of polynomials with \( Q_0(x) := 1 \) and \( Q_n := r^{-n}a_1a_2\cdots a_{n-1}C_n \) for all \( n \in \mathbb{N} \). Using (5) we obtain
\[
2xQ_1(x) = Q_2(x) + \frac{c_1}{r^2}Q_0(x),
\]
\[
2xQ_n(x) = Q_{n+1}(x) + Q_{n-1}(x), \quad n \geq 2.
\]
From \([1, (4.28)-(4.30)]\) we deduce for all \( \theta \in \mathbb{C} \setminus \{0\} \):
\[
Q_n(\cos \theta) = \sin((n+1)\theta) + \frac{k}{r^2} \sin(n\theta) - \frac{r^{-n}}{2^n} \sin((n-1)\theta) / \sin(\theta).
\]
Going back to the \( C_n \) this equation yields the desired representation.

To obtain the second claim we write
\[
B_n := \frac{b(a-1)}{a(ab-a-b)} \left( 1 - \frac{1}{(a-1)^n(b-1)^n} \right).
\]
Using the explicit form of \( m_1 \) given in Section 4 this yields
\[
e^{-im_1(x)x} = \cos(nx) \cos(B_nx) + \sin(nx) \sin(B_nx) - i[\sin(nx) \cos(B_nx) - \cos(nx) \sin(B_nx)].
\]
Since for small \( x \) we have \( \sin(x) = x + O(x^2) \) and \( \cos(x) = 1 + O(x^2) \) and since \( B_n \) is bounded a longer but straightforward computation shows for small \( x \):
\[
e^{-im_1(x)x} = B_nx \sin(nx) + O(x^2) - i[\sin(nx) - B_nx \cos(nx) + O(x^2)],
\]
\[
C_n(\cos(i\vartheta_0 + x)) = A_nx \sin(nx) + O(x^2) + i[B_nx \cos(nx) - \sin(nx) + O(x^2)]
\]
with
\[
A_n = \frac{b(a-1)}{a(ab-a-b)} \left( 1 + \frac{1}{(a-1)^n(b-1)^n} \right).
\]
Hence we have
\[
\text{Re}[C_n(\cos(i\vartheta_0 + x)) - e^{-im_1(x)x}] = x(A_n - B_n) \sin(nx) + O(x^2)
\]
with
\[
A_n - B_n = \frac{2b(a-1)}{a(ab-a-b)} \frac{1}{(a-1)^n(b-1)^n}.
\]
Therefore using \( |\sin(x)| \leq |x| \) for all \( x \in \mathbb{R} \) we obtain
\[
|\text{Re}[C_n(\cos(i\vartheta_0 + x)) - e^{-im_1(x)x}])| \leq \frac{2b(a-1)}{a(ab-a-b)} \frac{n}{(a-1)^n(b-1)^n} x^2 + O(x^2),
\]
where the rest \( O(x^2) \) does not depend on \( n \). Since \( a > \gamma > 0 \) either \( a \) or \( b \) has to be greater than 2. Therefore \( (a-1)(b-1) > 1 \) and
\[
\lim_{n \to \infty} \frac{2b(a-1)}{a(ab-a-b)} \frac{n}{(a-1)^n(b-1)^n} = 0.
\]
Hence

\[ |\text{Re}[C_n(\cos(i\vartheta_0 + x)) - e^{-im_1(x)}]| = O(x^2). \]

An analogous argument for the imaginary part of \( C_n(\cos(i\vartheta_0 - x)) - e^{im_1(x)} \) shows

\[ |\text{Im}[C_n(\cos(i\vartheta_0 + x)) - e^{-im_1(x)}]| = O(x^2), \]

where the rest \( O(x^2) \) does not depend on \( n \).

For the proof of the main theorem we need the following two lemmata.

**Lemma 6.1.** Let \((P_n)_{n \in \mathbb{N}_+}\) be a sequence of orthogonal polynomials as in Section 2. Then \( |P_n(\cos(i\vartheta_0 - \lambda))| \leq 1 \) for all \( \lambda \in \mathbb{R} \) and for a probability measure \( \nu \) on \( \mathbb{N}_0 \) the function \( g: \mathbb{R} \to \mathbb{C}, \lambda \mapsto \sum_{k=0}^{\infty} P_k(\cos(i\vartheta_0 - \lambda))\nu(\{k\}) \) is well defined. If we assume that \( \sum_{k=0}^{\infty} k^\epsilon\nu(\{k\}) < \infty \) for some \( n \in \mathbb{N}_0 \), then \( g \) is \( n \)-times differentiable and \( g^{(n)}(\lambda) = \sum_{k=0}^{\infty} (d/d\lambda)^n P_k(\cos(i\vartheta_0 - \lambda))\nu(\{k\}) \) holds. Furthermore, if \((S_n)_{n \in \mathbb{N}_+}\) is a stationary Markov chain then

\[ E(P_n(\cos(i\vartheta_0 - \lambda))) = E(P_k(\cos(i\vartheta_0 - \lambda)))^n \]

for all \( \lambda \in \mathbb{R} \).

The proof is given in [9].

**Lemma 6.2.** Let \( r \geq 1 \) be a real number and let \((P_n)_{n \in \mathbb{N}_+}\) be a sequence of polynomials as in Section 2 satisfying

\[ \sup_{k \in \mathbb{N}_0} |P_k(\cos(i\vartheta_0 - x)) - e^{im_1(x)}| = O(x^r) \]

for \( |x| \to 0 \). For \( \sigma > 0 \) define for all \( t \in \mathbb{R} \):

\[ h_n(t) := \left| E\left[ e^{im_1(S_n)/\sqrt{n}} - P_{S_n}\left( \cos\left( t\vartheta_0 - \frac{t}{\sqrt{n}} \right) \right) \right] \right|. \]

If \((K_n)_{n \in \mathbb{N}_+}\) is any sequence of real numbers with \( K_n \to \infty \) for \( n \to \infty \) then for all \( d > 0 \) there exists a \( C > 0 \) such that for all \( n \in \mathbb{N} \) and all \( t \in \mathbb{R} \) satisfying \( |t| \leq dK_n \) the following holds:

\[ h_n(t) \leq C |t|^r n^{-1/2}. \]

Especially for the Bernstein–Cartier polynomials we have \( h_n(t) \leq C |t| n^{-2/3} \) for \( |t| \leq dn^{-1/3} \).

**Proof.** Using the assumption \( \sup_{k \in \mathbb{N}_0} |P_k(\cos(i\vartheta_0 - x)) - e^{im_1(x)}| = O(x^r) \) there exists a \( C_1 > 0 \) such that

\[ h_n(t) \leq E\left| e^{im_1(S_n)/\sqrt{n}} - P_{S_n}\left( \cos\left( t\vartheta_0 - \frac{t}{\sqrt{n}} \right) \right) \right| \]

\[ \leq \sup_{k \in \mathbb{N}_0} |P_k(\cos(i\vartheta_0 - x)) - e^{im_1(x)}| \leq C_1 |t|^r n^{-1/2} . \]
Proof of Theorem 5.2. For all $t \in \mathbb{R}$ set $\varphi_n(t) := \mathbb{E}( e^{i(t(\xi_n) - n\theta)/\sqrt{n}} )$ and $\psi(t) := e^{-t^2/2}$. Applying Lemma 6.1 $g(x) := \mathbb{E}[P_{\xi}(\cos(i\theta_0 - x))] e^{iux} (x \in \mathbb{R})$ is three times differentiable and we get $g(x) = 1 - 1_2 \sigma^2 x^2 + O(x^3)$. For sufficiently small $K_1 > 0$ and $|x| \leq K_1$ we have

$$ |g(x)| \leq 1 - 1_2 \sigma^2 x^2 + |O(x^3)| \leq 1 - 1_4 \sigma^2 x^2. $$

Together with $1 - 1_4 \sigma^2 x^2 \leq e^{-\frac{\sigma^2}{2} x^2}$ for $|x| \leq K_1$ it follows that $|g(x)| \leq e^{-\frac{\sigma^2}{2} x^2}$ for $|x| \leq K_1$. From this last inequality and the mean value theorem for differentiation in Banach spaces (see [5, Satz 175.3]) we infer for $n \in \mathbb{N}$ and $t \leq K_1 \sigma \sqrt{n}$:

$$ |g\left(\frac{t}{\sigma \sqrt{n}}\right) - e^{-\frac{\sigma^2}{2} t^2}| \leq n \left|g\left(\frac{t}{\sigma \sqrt{n}}\right) - e^{-\frac{\sigma^2}{2} t^2}\right| \left|\max\left(g\left(\frac{t}{\sigma \sqrt{n}}, e^{-\frac{\sigma^2}{2} t^2}\right)\right)^{-1}ight| \leq n \left|g\left(\frac{t}{\sigma \sqrt{n}}\right) - e^{-\frac{\sigma^2}{2} t^2}\right| e^{-\frac{\sigma^2}{2} (n-1) t^2}. $$

Using $|g(t/\sigma \sqrt{n}) - e^{-\frac{\sigma^2}{2} t^2}| = O(t^3/\sqrt{n})$ the above inequality implies

$$ |g\left(\frac{t}{\sigma \sqrt{n}}\right) - e^{-\frac{\sigma^2}{2} t^2}| \leq K_2 \left|\frac{t}{\sqrt{n}}\right|^3 e^{-\frac{\sigma^2}{2} (n-1) t^2} \leq K_2 \left|\frac{t}{\sqrt{n}}\right|^3 e^{-\frac{\sigma^2}{2} t^2} \tag{6} $$

for a suitable $K_2 > 0$. Since $\mathbb{E}[P_{\xi}(\cos(i\theta_0 - x))] = \mathbb{E}[P_{\xi}(\cos(i\theta_0 - x))^n]$ (see Lemma 6.1) for all $n \in \mathbb{N}$ and $t \leq K_1 \sigma \sqrt{n}$ the following is valid using (6):

$$ |\varphi_n(t) - \psi(t)| \leq \left|\mathbb{E}\left[e^{i(t(\xi_0) - n\theta)/\sigma \sqrt{n}}\right] - \mathbb{E}\left[P_{\xi}\left(\cos\left(\frac{t}{\sigma \sqrt{n}}\right)\right)\right] e^{\frac{n\mu t}{\sigma \sqrt{n}}} \right| + \left|\mathbb{E}\left[P_{\xi}\left(\cos\left(\frac{t}{\sigma \sqrt{n}}\right)\right)\right] e^{\frac{n\mu t}{\sigma \sqrt{n}}} - e^{-t^2/2}\right| $$

$$ = h_n(t) + \left|g\left(\frac{t}{\sigma \sqrt{n}}\right) - e^{-\frac{\sigma^2}{2} t^2}\right| \leq h_n(t) + K_2 \left|\frac{t}{\sqrt{n}}\right|^3 e^{-\frac{\sigma^2}{2} t^2}. \tag{7} $$

By [3, p. 538] there exists a constant $K_3 > 0$ such that for all $K > 0$:

$$ \sup_{x \in \mathbb{R}} |F_n(x) - G(x)| \leq K_3 \int_{-K}^K \left|\frac{\varphi_n(t) - \psi(t)}{t}\right| dt + \frac{1}{K}. $$
Setting \( K := K_n \) with \( K_n \rightarrow \infty \) for \( n \rightarrow \infty \) and \( K_n \leq K_1 \sigma \sqrt{n} \) one obtains from Lemma 6.2, Eq. (7) and the last inequality:

\[
sup_{x \in \mathbb{R}} |F_n(x) - G(x)| \leq K_4 \left( \int_{-K_n}^{K_n} \frac{h_n(t)}{|t|} + \frac{|t|^2}{\sqrt{n}} e^{-\frac{\sigma^2}{8} t^2} dt + \frac{1}{K_n} \right)
\]

\[
\leq K_5 \left( \frac{K_n}{n^{3/2}} \int_{-K_n}^{K_n} dt + \frac{1}{K_n} + \int_{-K_n}^{K_n} \frac{t^2}{\sqrt{n}} e^{-\frac{\sigma^2}{8} t^2} dt \right)
\]

\[
= K_5 \left( \frac{2K_n^r}{n^{r/2}} + \frac{1}{K_n} + \frac{2}{\sqrt{n}} \int_{0}^{K_n} t^2 e^{-\frac{\sigma^2}{8} t^2} dt \right)
\]

for suitably chosen constants \( K_4, K_5 > 0 \). Together with

\[
\frac{2}{\sqrt{n}} \int_{0}^{K_n} t^2 e^{-\frac{\sigma^2}{8} t^2} dt \leq \frac{2}{\sqrt{n}} \int_{0}^{\infty} t^2 e^{-\frac{\sigma^2}{8} t^2} dt = \frac{\Gamma(3/2)8^{3/2}}{\sqrt{n}}
\]

it follows from the above inequality:

\[
sup_{x \in \mathbb{R}} |F_n(x) - G(x)| \leq K_6 \left( \frac{K_n^r}{n^{r/2}} + \frac{1}{K_n} + \frac{1}{\sqrt{n}} \right)
\]

for some \( K_6 > 0 \). To make the right hand side tend to 0 as fast as possible it is easy to see that one has to choose \( K_n \) such that \( K_n^r/n^{r/2} = O(1/K_n) \). This is true if \( K_n = O(n^{r(2r+2)}) \). Because of \( r/(2r + 2) \leq 1/2 \) it is true that \( K_n \leq K_1 \sigma \sqrt{n} \) if we set \( K_n := K_1 \sigma n^{r(2r+2)} \). Together with (8) this shows sup_{x \in \mathbb{R}} |F_n(x) - G(x)| = O(n^{-r(2r+2)}). Using Proposition 5.1 we obtain the result for the Cartier polynomials with \( r = 2 \).

References


