Rates of Convergence for Products of Random Stochastic 2×2 Matrices

Ralph Neininger
Institut für Mathematische Stochastik
Universität Freiburg
Eckerstr. 1
79104 Freiburg
Germany

May 17, 2001

Abstract

Products of independent identically distributed random stochastic 2×2 matrices are known to converge in distribution under a trivial condition. Rates for this convergence are estimated in terms of the minimal L_p-metrics and the Kolmogorov metric and applications to convergence rates of related interval splitting procedures are discussed.

AMS subject classifications. Primary: 60F05; secondary: 60D05, 60E05.

Key words. Random Stochastic Matrix; Rate of Convergence; Kolmogorov metric; Random Walk; Interval Splitting.

1 Introduction and Main Result

Let $\{(V_i, W_i) : i \in \mathbb{N}\}$ be an independent family of random vectors on the unit square $[0,1]^2$ with common joint distribution μ. Let T_i denote the random stochastic 2×2 matrices

$$T_i := \begin{bmatrix} 1 - V_i & V_i \\ 1 - W_i & W_i \end{bmatrix}, \quad i \geq 1,$$

and

$$\begin{bmatrix} 1 - A_n & A_n \\ 1 - B_n & B_n \end{bmatrix} := T_n \cdots T_1, \quad n \geq 1. \tag{1}$$

It is well-known that (A_n, B_n) converges weakly if and only if μ is not concentrated on $\{(0,1),(1,0)\}$ and that the limit is concentrated on the diagonal $\{(x,x) \in [0,1]^2 : 0 \leq x \leq 1\}$, i.e. is of the form $\mathcal{L}(Y,Y)$, where the distribution function of Y can be characterized as the solution of an integral equation; see Rosenblatt [9], Sun [10].

The aim of this paper is to derive (geometric) rates of convergence for the distributions of (A_n, B_n) and to discuss applications to interval splitting problems.

Let $\mathcal{M}^d$ denote the space of probability measures on $\mathbb{R}^d$. The minimal L_p-metric ℓ_p is defined on the subspace $\mathcal{M}_p^d \subset \mathcal{M}^d$ of measures with finite p-th moment for $p \geq 1$ by

$$\ell_p(\lambda, \nu) := \inf\{\|X - Z\|_p : \mathcal{L}(X) = \lambda, \mathcal{L}(Z) = \nu\}, \quad \lambda, \nu \in \mathcal{M}_p^d, \tag{2}$$

where $\|\cdot\|_p$ denotes the L_p-norm; the notation $\ell_p(X, Z) := \ell_p(\mathcal{L}(X), \mathcal{L}(Z))$ is used as well.

Theorem 1.1 *Let (A_n, B_n) be given as in (1) with a distribution $\mu = \mathcal{L}(V, W)$ on $[0,1]^2$ not being concentrated on $\{(0,1),(1,0)\}$. Then it holds for all $p \geq 1$ and $n \in \mathbb{N}$*

$$\ell_p((A_n, B_n), (Y, Y)) \leq \|(Y, 1-Y)\|_p \, R^n$$

and for the marginals

$$\ell_p(A_n, Y) \leq \|Y\|_p \, R^n, \quad \ell_p(B_n, Y) \leq \|1 - Y\|_p \, R^n,$$

with $R := \|W - V\|_p < 1$ and $\mathcal{L}(Y)$ being the unique fixed-point of the map

$$T : \mathcal{M}^1 \to \mathcal{M}^1, \quad \nu \mapsto \mathcal{L}((W - V)Z + V), \tag{3}$$

where (V, W) and Z are independent and $\mathcal{L}(Z) = \nu$.

Since ℓ_p-convergence is equivalent to weak convergence plus convergence of the pth absolute moments (see Bickel and Freedman [2]) Theorem 1.1 implies the result stated above (the 'only if' part there being trivial) and endows it with a rate of convergence. The characterization of the limit distribution as the fixed-point of T is well-known. Other rates of convergence in terms of the random variables $N_\delta = \max\{n \in \mathbb{N} : |A_n - B_n| \geq \delta\}$ were discussed by van Assche [1].

From (1) the distributions $\mathcal{L}(A_n, B_n)$ are obtained by iterating the map

$$S_1 : \mathcal{M}^2 \to \mathcal{M}^2, \quad \nu \mapsto \mathcal{L}\left(\begin{bmatrix} 1-V & V \\ 1-W & W \end{bmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \right), \tag{4}$$

where (V, W) and $Z = (Z_1, Z_2)$ are independent with $\mathcal{L}(Z) = \nu$. More formally, it is $\mathcal{L}(A_n, B_n) = S_1^n(\delta_{(0,1)}) := S_1 \circ \cdots \circ S_1(\delta_{(0,1)})$, the n-th iteration of S_1 applied to the Dirac measure $\delta_{(0,1)}$ in $(0,1)$. The present approach is based on the fact (Lemma 3.1) that the $\mathcal{L}(A_n, B_n)$ can also be obtained as the iteration $S_2^n(\delta_{(0,1)})$ with

$$S_2 : \mathcal{M}^2 \to \mathcal{M}^2, \quad \nu \mapsto \mathcal{L}\left((W - V) \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} + \begin{pmatrix} V \\ V \end{pmatrix} \right), \tag{5}$$

where (V, W) and $Z = (Z_1, Z_2)$ are independent with $\mathcal{L}(Z) = \nu$. The key point is that the map S_2 is a contraction on the complete metric space $(\mathcal{M}_p^2, \ell_p)$ under the conditions of Theorem 1.1, so that its proof breaks essentially down to the application of Banach's fixed-point theorem. Clearly, S_1 cannot be a contraction on any $\mathcal{M}_p^2$ in any metric since the map has many fixed-points. For each random variable X, $\mathcal{L}(X, X)$ is a fixed-point of S_1.

As a standard metric for the quantification of weak convergence the Kolmogorov (or uniform) metric ϱ may be considered,

$$\varrho(\lambda, \nu) := \sup_{x \in \mathbb{R}} |F_\lambda(x) - F_\nu(x)|,$$

where F_λ, F_ν denote the distribution functions of one-dimensional probability measures λ, ν respectively; the notation $\varrho(X,Z) := \varrho(X,\mathcal{L}(Z)) := \varrho(\mathcal{L}(X),\mathcal{L}(Z))$ is also used. A change from the ℓ_p-metric to ϱ is provided by an inequality related to Markov's inequality,

$$\varrho(Y_n, Y) \le \ell_p^p(Y_n, Y)\xi^{-p} + \Delta_Y(\xi), \tag{6}$$

valid for all $p \ge 1$ and $\xi > 0$, where

$$\Delta_Y(\xi) := \sup_{x\in\mathbb{R}} |F(x+\xi) - F(x)| \tag{7}$$

denotes the modulus of continuity of Y, F being the distribution function of Y. Therefore the rates of Theorem 1.1 can be transposed into a rate for the Kolmogorov metric if the modulus of continuity of the limit $\mathcal{L}(Y)$ can be estimated. However, $\mathcal{L}(Y)$ is only known explicitly for a few choices of μ. For some μ related to interval splitting procedures we follow this line in the next section (Corollary 1 and 2) leading to explicit rates in the Kolmogorov metric. The last section contains the proofs.

2 Applications to interval splitting

Sequences of nested random intervals $([A_n, B_n])$ which are defined by $[A_0, B_0] := [0,1]$ and some randomized recursive procedure may be covered by a formulation as in (1). Then one is interested in the limit Y to which the intervals shrink almost surely. Theorem 1.1 implies corresponding rates of convergence for the end-points A_n, B_n. We give two examples for such an interval splitting procedure and obtain rates of convergence. When the modulus of continuity of the limit Y is available the rates can be given as well in the Kolmogorov metric.

2.1 Chen, Goodman, and Zame's splitting procedure

Chen, Goodman, and Zame [3] and Chen, Lin, and Zame [4] considered the following recursive interval splitting procedure: Fix $r \in [0,1]$ and set $[A_0, B_0] := [0,1]$. If $[A_n, B_n]$ is already defined, then split $[A_n, B_n]$ by an independent and uniformly on $[A_n, B_n]$ distributed random variable X and choose independently the larger of the two subintervals $[A_n, X], [X, B_n]$ with probability r to be $[A_{n+1}, B_{n+1}]$, otherwise the smaller one.

In the papers mentioned it is proved that $([A_n, B_n])$ shrinks to a limit Y almost surely, where Y has the beta$(2,2)$ distribution if $r = 1$ and the arcsin$(=$ beta$(1/2,1/2))$ distribution if $r = 1/2$ (see also Devroye, Letac, and Seshadri [5]).

For the analysis of this interval splitting procedure it is convenient to represent a uniform $[0,1]$ random variable in the form

$$G\frac{1+U}{2} + (1-G)\frac{1-U}{2}$$

with independent G, U with $\mathcal{L}(U) = \mathrm{unif}[0,1]$ and $\mathcal{L}(G) = \mathrm{B}(1,1/2)$, the Bernoulli distribution with probability $1/2$ on the point 1. Using such a representation for the splitting random variable in the definition of $[A_{n+1}, B_{n+1}]$ one may find that (A_n, B_n) is given by (1) choosing $\mu = \mathcal{L}(V,W)$ with

$$V = G'(1-G)\frac{1-U}{2} + (1-G')G\frac{1+U}{2}, \tag{8}$$

$$W = 1 - G'G\frac{1-U}{2} - (1-G')(1-G)\frac{1+U}{2}, \tag{9}$$

where G, G', U are independent with $\mathcal{L}(G') = \mathrm{B}(1, r), \mathcal{L}(G) = \mathrm{B}(1, 1/2), \mathcal{L}(U) = \mathrm{unif}[0, 1]$. From Theorem 1.1 we obtain

Corollary 2.1 *Let $([A_n, B_n])$ be the interval splitting procedure of Chen, Goodman and Zame with $r \in [0, 1]$. Then it holds for all $p \geq 1$ and $n \in \mathbb{N}_0$*

$$\ell_p(A_n, Y) = \ell_p(B_n, Y) \leq \|Y\|_p \left(\frac{2(r + (1 - 2r)(1/2)^{p+1})}{p + 1} \right)^{n/p},$$

where $\mathcal{L}(Y)$ is the unique fixed-point of (3) with (V, W) given by (8), (9).
In the case $r = 1$ it holds

$$\varrho(A_n, \mathrm{beta}(2, 2)) = \varrho(B_n, \mathrm{beta}(2, 2)) \leq 1.5661 \cdot (0.8268)^n.$$

In the case $r = 1/2$ it holds

$$\varrho(A_n, \arcsin) = \varrho(B_n, \arcsin) \leq 1.6321 \cdot (0.793)^n.$$

The case $r = 0$ leads to a distribution with a density, which is not infinitely differentiable on a dense subset of $[0, 1]$. Properties of this distribution were studied in Chen, Goodman, and Zame [3] and (in a more general situation) by Herz [6]. For all $r \in [0, 1]$ the limit Y has the representation

$$Y \stackrel{\mathcal{D}}{=} \sum_{j=0}^{\infty} \left(\left(G_j (1 - G'_j) \frac{1 + U_j}{2} + (1 - G_j) G'_j \frac{1 - U_j}{2} \right) \prod_{k=0}^{j-1} \frac{1 + S_k U_k}{2} \right)$$

with the family $\{G_n, G'_n, U_n : n \in \mathbb{N}_0\}$ of independent random variables with $\mathcal{L}(G'_n) = \mathrm{B}(1, r), \mathcal{L}(G_n) = \mathrm{B}(1, 1/2), \mathcal{L}(U_n) = \mathrm{unif}[0, 1]$ and the random sign $S_k := 2G'_k - 1$. This follows since the series converges almost surely, e.g., by the Borel-Cantelli Lemma, and satisfies the fixed-point relation. For the case $r = 1$ Chen et al. [3] gave this representation for a $\mathrm{beta}(2, 2)$ distributed random variable. The case $r = 1/2$ gives a corresponding representation for the arcsin distribution.

In Herz [6] generalizations of the cases $r \in \{0, 1\}$ of this splitting procedure were investigated, which could also be endowed with a rate of convergence by the present approach.

2.2 Kennedy's splitting procedure

The analysis of a randomized algorithm for locating local maxima led Kennedy [7] to the following splitting procedure: Let $k \in \mathbb{N}$, $k \geq 2$ and $r, s, t \in [0, 1]$ with $r + s + t = 1$ be given and define $[A_0, B_0] := [0, 1]$. If $[A_n, B_n]$ is already defined, then draw $X_1, \ldots, X_k$ independently and uniformly from $[A_n, B_n]$ and denote $C_n := \min_{1 \leq i \leq k} X_i$ and $D_n := \max_{1 \leq i \leq k} X_i$. Then $[A_{n+1}, B_{n+1}]$ is independently chosen from the three intervals $[C_n, B_n]$, $[A_n, D_n]$, and $[C_n, D_n]$ with probabilities r, s, t respectively.

Kennedy showed that $([A_n, B_n])$ shrinks almost surely to a limit with $\mathrm{beta}(k(r + t), k(s + t))$ distribution. Let G denote a random variable with

$$\mathbb{P}(G = 0) = r, \quad \mathbb{P}(G = 1) = s, \quad \mathbb{P}(G = 2) = t.$$

Then (A_n, B_n) is given by (1) choosing $\mu = \mathcal{L}(V, W)$ with

$$V = \mathbf{1}[G \in \{0, 2\}] \min_{1 \leq i \leq k} U_i, \quad W = 1 - \mathbf{1}[G \in \{1, 2\}](1 - \max_{1 \leq i \leq k} U_i), \tag{10}$$

where $U_1, \ldots, U_k, G$ are independent, $\mathcal{L}(U_i) = \mathrm{unif}[0, 1]$ for $i = 1 \ldots, k$ and $\mathbf{1}[A]$ denotes the indicator function of a set A. From Theorem 1.1 we obtain:

Corollary 2.2 *Let* $([A_n, B_n])$ *be Kennedy's interval splitting procedure with* $k \geq 2$ *and* $r,\ s,\ t \in [0,1]$ *with* $r+s+t=1$*. Then it holds for all* $p \geq 1$ *and* $n \in \mathbb{N}_0$

$$\begin{aligned}
&\ell_p(A_n, \mathrm{beta}(k(r+t), k(s+t)) \\
&\quad \leq \left[\frac{B(k(r+t)+p, k(s+t))}{B(k(r+t), k(s+t))}\right]^{1/p} \left(\frac{k}{k+p}\left(r+s+t\frac{k-1}{k-1+p}\right)\right)^{n/p}, \\
&\ell_p(B_n, \mathrm{beta}(k(r+t), k(s+t)) \\
&\quad \leq \left[\frac{B(k(r+t), k(s+t)+p)}{B(k(r+t), k(s+t))}\right]^{1/p} \left(\frac{k}{k+p}\left(r+s+t\frac{k-1}{k-1+p}\right)\right)^{n/p},
\end{aligned}$$

where $B(\,\cdot\,,\cdot\,)$ *denotes the Eulerian beta integral.*

Transformations to rates in the Kolmogorov metric can be given for this splitting procedure as well, since the modulus of continuity of the beta distributions can be estimated. However, numerical solutions required to determine optimal values for p can only be obtained if the parameters k, r, s, t are given explicitly. For example for the case $k=2, t=1, r=s=0$, which occurred also as a special case in van Assche [1] and Letac and Scarsini [8], we obtain

$$\varrho(A_n, \mathrm{beta}(2,2)) = \varrho(B_n, \mathrm{beta}(2,2)) \leq 1.6643 \cdot (0.5503)^n. \tag{11}$$

3 Proofs

In this section we use throughout the representation $\mu = \mathcal{L}(V, W)$ for the measure μ in the definition (1) as well as the family $\{(V_i, W_i) : i \in \mathbb{N}\}$ occurring there.

Lemma 3.1 *With* (A_n, B_n) *given by (1) it holds* $\mathcal{L}(A_{n+1}, B_{n+1}) = S_2(\mathcal{L}(A_n, B_n))$ *for all* $n \geq 1$*, where* S_2 *is given by (5).*

Proof: Using (1) we obtain

$$(A_1, B_1) = (V_1, W_1), \quad (A_2, B_2) = (V_1 + V_2(W_1 - V_1), V_1 + W_2(W_1 - V_1)),$$

and by induction for $n \geq 1$

$$\begin{aligned}
A_n &= \sum_{i=1}^{n} \left(V_i \prod_{k=1}^{i-1}(W_k - V_k)\right), \qquad (12)\\
B_n &= \sum_{i=1}^{n-1} \left(V_i \prod_{k=1}^{i-1}(W_k - V_k)\right) + W_n \prod_{k=1}^{n-1}(W_k - V_k),
\end{aligned}$$

where empty sums and products are defined to be 0 and 1 respectively. Thus, with (V, W) being independent of $\{(V_i, W_i) : i \in \mathbb{N}\}$ we obtain the recursion stated. ■

Lemma 3.2 *The restriction* $S_2 : \mathcal{M}_p^2 \to \mathcal{M}_p^2$ *of* S_2 *given in (5) is Lipschitz continuous w.r.t.* ℓ_p *for all* $p \geq 1$*:*

$$\ell_p(S_2(\lambda), S_2(\nu)) \leq \|W - V\|_p \, \ell_p(\lambda, \nu) \tag{13}$$

for all $\lambda,\ \nu \in \mathcal{M}_p^2$*.*

Proof: Let $\nu \in \mathcal{M}_p^2$ and (V, W), Z be independent with $\mathcal{L}(Z) = \nu$. Since (V, W) has also a finite p-th moment and by independence also $S_2(\nu)$ has a p-th moment, so $S_2 : \mathcal{M}_p^2 \to \mathcal{M}_p^2$ is well-defined. A property of the ℓ_p-metric is that for all λ, $\nu \in \mathcal{M}_p^2$ there exist random variables Z, Z' with $\mathcal{L}(Z) = \lambda$, $\mathcal{L}(Z') = \nu$, and $\ell_p(\lambda, \nu) = \|Z - Z'\|_p$, i.e. the infimum in (2) is in fact a minimum (see Bickel and Freedman [2]). We can choose (Z, Z') to be independent of (V, W). Then it follows

$$\begin{aligned} \ell_p(S_2(\lambda), S_2(\nu)) &\leq \|(W - V)(Z - Z')\|_p \\ &= \|W - V\|_p \|Z - Z'\|_p \\ &= \|W - V\|_p \ell_p(\lambda, \nu), \end{aligned}$$

valid for all λ, $\nu \in \mathcal{M}_p^2$. ■

Proof of Theorem 1.1: The assumption $\mu(\{(0, 1), (1, 0)\}) < 1$ implies $R := \|W - V\|_p < 1$, hence by Lemma 3.2 the restriction of S_2 to $\mathcal{M}_p^2$ is a contraction. Since $(\mathcal{M}_p^2, \ell_p)$ is a complete metric space (see Bickel and Freedman [2]), Banach's fixed-point theorem yields a unique fixed-point $\mathcal{L}(Y_1, Y_2) \in \mathcal{M}_p^2$ for the map S_2 on $\mathcal{M}_p^2$ and that $S_2^n(\nu)$ converges to $\mathcal{L}(Y_1, Y_2)$ in ℓ_p for every $\nu \in \mathcal{M}_p^2$. Choosing $\nu = \mu$ we obtain with $\mathcal{L}(A_1, B_1) = \mu$ and (by Lemma 3.1)

$$\mathcal{L}(A_n, B_n) = S_2^{n-1}(\mu), \quad n \in \mathbb{N},$$

that (A_n, B_n) converges in ℓ_p to the fixed-point $\mathcal{L}(Y_1, Y_2)$.

On the other hand it is $\mathbb{E}\,|A_n - B_n| = \mathbb{E}\,|\prod_{k=1}^n (W_k - V_k)| \to 0$ at an exponential rate, thus by the Borel-Cantelli Lemma it holds $|A_n - B_n| \to 0$ almost surely. Since the intervals $[A_n, B_n]$ resp. $[B_n, A_n]$ are nested it follows that (A_n), (B_n) converge to the same limit almost surely; this implies $Y_1 = Y_2$ almost surely, thus the fixed-point is of the form $\mathcal{L}(Y, Y)$.

By the definitions of S_2 and T in (3) and (5) it follows that $\mathcal{L}(Y)$ is a fixed-point of T. It holds $\mathbb{E}\ln|W - V| \leq \ln \mathbb{E}\,|W - V| \in [-\infty, 0)$, thus Lemma 1.4 (a) and Theorem 1.5 (i) in Vervaat [11] imply that T has a unique fixed-point in $\mathcal{M}^1$. Therefore $\mathcal{L}(Y, Y)$ and $\mathcal{L}(Y)$ are the unique fixed-points of S_2 and T on the whole spaces $\mathcal{M}^2$ and $\mathcal{M}^1$ respectively.

With the Dirac measure $\delta_{(0,1)}$ in $(0, 1)$ it holds $S_2(\delta_{(0,1)}) = \mu$, hence iterated application of Lemmas 3.1 and 3.2 implies

$$\begin{aligned} \ell_p((A_n, B_n), (Y, Y)) &= \ell_p(S_2(A_{n-1}, B_{n-1}), S_2(Y, Y)) \\ &\leq R\,\ell_p((A_{n-1}, B_{n-1}), (Y, Y)) \\ &\leq \ell_p(\mu, (Y, Y))\, R^{n-1} \\ &= \ell_p(S_2(\delta_{(0,1)}), S_2(Y, Y))\, R^{n-1} \\ &\leq \ell_p(\delta_{(0,1)}, (Y, Y))\, R^n \\ &= \|(Y, 1 - Y)\|_p\, R^n. \end{aligned}$$

The estimates for the marginals in Theorem 1.1 can be deduced the same way using the map T defined in (3) instead of S_2. ■

Note that we could have also used the representations

$$\mathcal{L}(Y) = \mathcal{L}\left(\sum_{i=1}^{\infty}\left(V_i \prod_{k=1}^{i-1}(W_k - V_k)\right)\right),$$

and (12) for $\mathcal{L}(A_n, B_n)$ to obtain an estimate for $\ell_p((A_n, B_n), (Y, Y))$. This would have led to a worse rate of convergence.

Proof of Corollary 2.1: A direct computation shows that V, W given in (8), (9) satisfy

$$\|W - V\|_p = \left(\frac{2(r + (1-2r)(1/2)^{p+1})}{p+1}\right)^{1/p}, \tag{14}$$

which by Theorem 1.1 implies the estimate for the ℓ_p metric.

In the case $r = 1$ it is $\mathcal{L}(Y) = \mathrm{beta}(2, 2)$, thus Δ_Y defined in (7) is given by $\Delta_Y(\xi) = (3\xi - \xi^3)/2$ for $\xi \in [0, 1]$, in particular it holds $\Delta_Y(\xi) \leq (3/2)\xi$ for $\xi > 0$. This implies with $r = 1$ in (14) and (6)

$$\rho(A_n, \mathrm{beta}(2,2)) \leq \inf_{p \geq 1} \inf_{\xi > 0} c_p^p \left(\frac{2(1 - (1/2)^{p+1})}{p+1}\right)^n \xi^{-p} + \frac{3}{2}\xi$$

with

$$c_p^p := \|Y\|_p^p = \frac{6}{(p+2)(p+3)}.$$

To optimize the bound we choose

$$\xi = \xi_n = \alpha \left(\frac{2(1 - (1/2)^{p+1})}{p+1}\right)^{n/(p+1)}$$

with $\alpha > 0$. This implies

$$\begin{aligned}
\rho(Y_n, \mathrm{beta}(2,2)) &\leq \inf_{p \geq 1} \inf_{\alpha > 0} \left(\left(\frac{c_p}{\alpha}\right)^p + \frac{3}{2}\alpha\right) \left(\frac{2(1 - (1/2)^{p+1})}{p+1}\right)^{n/(p+1)} \\
&= \inf_{p \geq 1} \left(\left(\frac{c_p}{\alpha_p}\right)^p + \frac{3}{2}\alpha_p\right) \left(\frac{2(1 - (1/2)^{p+1})}{p+1}\right)^{n/(p+1)}
\end{aligned}$$

with

$$\alpha_p := \left(\frac{2pc_p^p}{3}\right)^{1/(p+1)}.$$

An approximation of the optimal choice of p is given by $p_0 = 3.4969$. This leads to the constants given in the corollary. By symmetry, $\varrho(B_n, \mathrm{beta}(2,2)) = \varrho(A_n, \mathrm{beta}(2,2))$.

For $r = 1/2$ the modulus of continuity Δ_Y of the arcsin distribution is given by $\Delta_Y(\xi) = (2/\pi)\arcsin(\sqrt{\xi})$ for $\xi \in [0, 1]$. Using $\arcsin(x) = \arctan(x/\sqrt{1 - x^2})$ for $|x| < 1$ and the power series representation of the arctan-function we estimate $\Delta_Y(\xi) \leq (2/\pi)\sqrt{\xi/(1 - \xi)}$ for $0 < \xi < 1$. Proceeding as in the case $r = 1$ with

$$\xi = \xi_n = \alpha \left(\frac{1}{p+1}\right)^{2n/(1+2p)}$$

we derive

$$\rho(A_n, \arcsin) \leq \inf_{p \geq 1} \inf_{0 < \alpha < 1} \left(\left(\frac{c_p}{\alpha}\right)^p + \frac{2}{\pi}\sqrt{\frac{\alpha}{1 - \alpha}}\right) \left(\frac{1}{p+1}\right)^{n/(1+2p)}$$

with

$$c_p^p := \|Y\|_p^p = \frac{B(p+1/2,1/2)}{\pi}.$$

With $p = 1.155$ and $\alpha = 0.601$ we obtain the constants given in the corollary. ∎

Proof of Corollary 2.2: With V, W given in (10) it holds

$$\begin{aligned} W - V &= \mathbf{1}[G=0](1 - \min_{1\le i\le k} U_i) + \mathbf{1}[G=1] \max_{1\le i\le k} U_i \\ &\quad + \mathbf{1}[G=2](\max_{1\le i\le k} U_i - \min_{1\le i\le k} U_i). \end{aligned}$$

Using $\mathcal{L}(\max U_i - \min U_i) = \text{beta}(k-1,2)$ and $\mathcal{L}(1 - \min U_i) = \mathcal{L}(\max U_i) = \text{beta}(k,1)$ leads to the value stated for $\|W - V\|_p$. Theorem 1.1 finishes the proof of Corollary 2.2. In the case $k = 2, t = 1, r = s = 0$ the transformation to the Kolmogorov metric can be done as in the case $r = 1$ in the proof of Corollary 2.1 leading to the numerical values in (11). ∎

Acknowledgment

I thank L. Rüschendorf for his support in the first stage of the paper, L. Devroye for pointing to some related literature, and the referee for clarifying remarks on the subject.

References

[1] v. Assche, W. (1986). Products of 2×2 stochastic matrices with random entries. *J. Appl. Probab.* **23**, 1019–1024.

[2] Bickel, P. J. and Freedman, P. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9**, 1196–1217.

[3] Chen, R., Goodman, R. and Zame, A. (1984). Limiting distributions of two random sequences. *J. Mult. Ana.* **14**, 221–230.

[4] Chen, R., Lin, E. and Zame, A. (1981). Another acr sine law. *Sankhyā Ser. A* **43**, 371–373.

[5] Devroye, L., Letac, G. and Seshadri, V. (1986). The limit behavior of an interval splitting scheme. *Statist. Probab. Lett.* **4**, 183–186.

[6] Herz, C. (1988). Splitting intervals. *Statist. Probab. Lett.* **7**, 3–7.

[7] Kennedy, D. P. (1988). A note on stochastic search methods for global optimization. *Adv. in Appl. Probab.* **20**, 476–478.

[8] Letac, G. and Scarsini, M. (1998). Random nested tetrahedra. *Adv. in Appl. Probab.* **30**, 619–627.

[9] Rosenblatt, M. (1964). Equicontinuous Markov operators. *Teor. Verojatnost. i Primenen.* **9**, 205–222.

[10] Sun, T.-C. (1975). Limits of convolutions of probability measures on the set of 2×2 stochastic matrices. *Bull. Inst. Math. Acad. Sinica* **3**, 235–248.

[11] Vervaat, W. (1979). On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. in Appl. Probab.* **11**, 750–783.