Communications of the Korean Statistical Society
2009, Vol. 16, No. 1, 157–162

The Uniform Law of Large Numbers for the Baker Transformation

Jongsig Bae[1,a], Changha Hwang[b], Jooyong Shim[c]

[a]Dept. of Mathematics, SungKyunKwan Univ.,
[b]Div. of Information and Computer Science, Dankook Univ.,
[c]Dept. of Applied Statistics, Catholic Univ. of Daegu

Abstract

The baker transformation is an ergodic transformation defined on the half open unit square. This paper considers the limiting behavior of the partial sum process of a martingale sequence constructed from the baker transformation. We get the uniform law of large numbers for the baker transformation.

Keywords: Baker transformation, law of large numbers, uniform law of large numbers, martingales.

1. Introduction

In this paper, we obtain uniform versions of the law of large numbers (LLN) for the process generated by the baker transformation.

In obtaining the uniform LLN, we observe the role played by bracketing of the indexed class of functions of the process and slightly modify the underlying metric. Then we employ the idea of DeHardt (1971) of the bracketing method to the process generated by the baker transformation.

Let X be a random variable defined on a probability space $(\Omega, \mathcal{T}, P)$ whose distribution function is F. Consider a sequence $\{X_i : i \geq 1\}$ of independent copies of X. Given a Borel measurable function $f : \mathbf{R} \to \mathbf{R}$, we see that $\{f(X_i) : i \geq 1\}$ forms a sequence of IID random variables that are more flexible in applications than the sequence $\{X_i : i \geq 1\}$. Consider a class $\mathcal{F}$ of real-valued Borel measurable functions defined on $\mathbf{R}$. Introduce the usual empirical distribution function F_n defined by

$$F_n(x) = n^{-1} \sum_{i=1}^{n} I_{\{X_i \leq x\}}, \qquad \text{for } x \in \mathbf{R}.$$

Define a function indexed integral process S_n by

$$S_n(f) = \int f(x)d(F_n - F)(x), \qquad \text{for } f \in \mathcal{F}. \tag{1.1}$$

Define a function indexed integral process S_n by

$$S_n(f, s) = \int f(x)d\left(F_{[ns]} - F\right)(x), \qquad \text{for } f \in \mathcal{F} \text{ and } s \in [0, 1], \tag{1.2}$$

[1] Corresponding author: Professor, Department of Mathematics, SungKyunKwan University, Suwon 440-746, Korea.
E-mail: jsbae@skku.edu

where $[x]$ is the integer part of x.

Developing a uniform LLN for the function-indexed process such as $S_n(f)$ in (1.1) is usually meant that $\sup_{f\in\mathcal{F}} |S_n(f)|$ converges to zero in a certain sense under a certain entropy condition on the class $\mathcal{F}$. The process is indexed by $\mathcal{F}$ and is considered as random elements in $B(\mathcal{F})$, the space of the bounded real-valued functions on $\mathcal{F}$, taken with the sup norm $\|\cdot\|_{\mathcal{F}}$. It is known that $(B(\mathcal{F}), \|\cdot\|_{\mathcal{F}})$ forms a Banach space.

Let us denote $A \otimes B$ to be the set $\{(a, b) : a \in A, b \in B\}$ of all ordered pairs (a, b). Write $\mathcal{S} := \mathcal{F} \otimes [0, 1]$. Developing a uniform LLN for the random fields such as $S_n(f, s)$ in (1.2) is usually meant that $\sup_{(f, s)\in\mathcal{S}} |S_n(f, s)|$ converges to zero in a certain sense under a certain entropy condition on the class $\mathcal{S}$. The process is indexed by $\mathcal{S}$ and is considered as random elements in $B(\mathcal{S})$, the space of the bounded real-valued functions on $\mathcal{S}$, taken with the sup norm $\|\cdot\|_{\mathcal{S}}$. The result may be used in the nonparametric statistical inference. See Van de Geer (2000) for applications.

We use the following definition of almost sure convergence and convergence in the mean. See, for a recent reference, Van der Vaart and Wellner (1996). See also Hoffmann-Jϕrgensen (1991).

Definition 1. *A sequence of $B(\mathcal{F})$-valued random functions $\{Y_n\}$ converges with probability* 1 *to a constant c if*

$$P^*\left\{\sup_{f\in\mathcal{F}} Y_n(f) \to c\right\} = P\left\{\sup_{f\in\mathcal{F}} Y_n(f)^* \to c\right\} = 1.$$

The sequence $\{Y_n\}$ converges in the mean to a constant c if

$$E^* \sup_{f\in\mathcal{F}} Y_n(f) = E \sup_{f\in\mathcal{F}} Y_n(f)^* \to c.$$

Here E^* denotes the upper expectation with respect to the outer probability P^* and $\sup_{f\in\mathcal{F}} Y_n(f)^*$ is the measurable cover function of $\sup_{f\in\mathcal{F}} Y_n(f)$.

Remark 1. If the measurability of the process Y_n is guaranteed then the convergence in the above definition boils down to the usual almost sure convergence and convergence in the mean.

In 1971, DeHardt (1971) obtained the uniform LLN for the sequence of IID random variables under bracketing entropy. DeHardt's result states that if $\mathcal{F}$ has a bracketing entropy then $\sup_{\mathcal{F}} |S_n| \to 0$ almost surely.

The aim of our work is to develop the LLN, the uniform LLN and the uniform LLN for the random fields for the process generated by the baker transformation by employing DeHardt's idea of bracketing method. Our results will be stated as the mean convergence as well as the almost sure convergence.

In Section 2, we introduce the baker transformation and states the main results.

In Section 3, we provide the proofs for the main results.

2. The Main Results

We illustrate the baker transformation. Let $\Omega = [0, 1) \times [0, 1)$ be the sample space, $\mathcal{F}$ be the Borel sets and P be the Lebsgue measure. The baker transformation on the half open unit square is defined by

$$\phi(x, y) \longrightarrow \left(2x, \frac{y}{2}\right), \quad 0 \le x < \frac{1}{2}$$

and

$$\phi(x, y) \longrightarrow \left(2x - 1, \frac{y+1}{2}\right), \quad \frac{1}{2} \leq x < 1.$$

The name comes from imagining the unit square to be bread dough that is stretched in the x-direction until it is twice as long and half as high and then cut along $x = 1$ to make two loaves.

We can think about $(\ldots, x_{-1}, x_0, x_1, \ldots) \in \{0, 1\}^Z$ as a point (x, y) in the half open unit square $[0, 1) \times [0, 1)$ by putting

$$x = \sum_{i=0}^{\infty} \frac{x_i}{2^{i+1}}$$

and

$$y = \sum_{i=1}^{\infty} \frac{x_{-i}}{2^i}.$$

It is known that the transformation is ergodic.

For $t \in [0, 1]$, we define

$$f_t(x, y) = 1_{[0,t]}(y), \quad 0 \leq x < \frac{1}{2}$$
$$= -1_{[0,t]}(y), \quad \frac{1}{2} \leq x < 1.$$

Consider the class of functions $\mathcal{G} = \{f_t(x, y)\}_{0 \leq t \leq 1}$. We denote $\phi^i(x, y) = (x_i, y_i)$, $(x, y) \in [0, 1) \times [0, 1)$, $i = 0, 1, \ldots$. Define

$$B_n(t) := B_n(f_t) = \frac{1}{n} \sum_{i=0}^{n-1} f_t(x_i, y_i), \quad f_t \in \mathcal{G}.$$

We are interested in the uniform limiting behavior of the process B_n.

We firstly state the following law of large numbers that will turn out to be a special case of an ergodic theorem. See Durrett (1991).

Theorem 1. *For each fixed $t \in [0, 1]$, as $n \to \infty$,*

$$B_n(t) \to 0$$

almost surely and in the mean.

We secondly state the following uniform law of large numbers for the process generated by the baker transformation.

Theorem 2. *As $n \to \infty$,*

$$\sup_{0 \leq t \leq 1} |B_n(t)| \to 0$$

almost surely and in the mean.

We generalize the above uniform law of large numbers to the context of random fields. Define

$$B_n(t, s) := B_n(f_t, s) = \frac{1}{n} \sum_{i=0}^{[(n-1)s]} f_t(x_i, y_i), \quad f_t \in \mathcal{G}, \ s \in [0, 1].$$

We thirdly state the following uniform law of large numbers for the process of random fields generated by the baker transformation.

Theorem 3. *As* $n \to \infty$,

$$\sup_{0 \le t \le 1} \sup_{0 \le s \le 1} |B_n(t, s)| \to 0$$

almost surely and in the mean.

3. A Setup of Martingale and the Proofs

In the proof of theorems we use the following setup of martingale difference. ($\Omega = S^Z$, $\mathcal{T} = \mathcal{B}^Z$, P) will be our basic probability space. We denote by T the left shift on Ω. We assume that P is invariant under T, *i.e.*, $PT^{-1} = P$ and that T is ergodic. We denote by $X = \ldots, X_{-1}, X_0, X_1, \ldots$ the coordinate maps on Ω. From our assumptions it follows that $\{X_i\}_{i \in Z}$ is a stationary and ergodic process. Next we define for each $i \in Z$ a σ-fields $M_i := \sigma(X_j : j \le i)$ and $H_i := \{f : \Omega \to R : f \in M_i \text{ and } f \in L^1(\Omega)\}$. We also denote for each $f \in L^1(\Omega)$, $E_{i-1}(f) := E(f \,|\, M_{i-1})$ and $H_0 \ominus H_{-1} := \{f \in H_0 : E(f \cdot g) = 0 \text{ for each } g \in H_{-1}\}$. Finally for every $f,\ g \in L^1(\Omega)$ we put

$$d(f, g) := \|f - g\| := E|f - g|.$$

Let us T^i denote the i^{th} iteration of T. If it were $\mathcal{F} \subseteq H_0 \ominus H_{-1}$ then from our setup it follows that for each fixed $f \in \mathcal{F}$, $\{f(T^i(X)),\ M_i\}$ is a stationary martingale difference sequence. We write $V_i := T^i(X)$ and $V := T^0(X)\ (= X)$. In that case, one can discuss the limiting behaviors of the process such as

$$S_n(f) = \frac{1}{n} \sum_{i=1}^{n} f(V_i) \quad \text{and} \quad S_n(f, t) = \frac{1}{n} \sum_{i=1}^{[nt]} f(V_i).$$

In order to measure the size of the function space, we define the following version of metric entropy with bracketing. See, for example, Van der Vaart and Wellner (1996) for the recent reference.

Definition 2. *Given two functions l and u, the bracket $[l, u]$ is the set of all functions f with $l \le f \le u$. An ϵ-bracket is a bracket $[l, u]$ with $\|u - l\| < \epsilon$. The bracketing number $N_{[\,]}(\epsilon) := N_{[\,]}(\epsilon, \mathcal{F}, d)$ is the minimum number of ϵ-brackets needed to cover $\mathcal{F}$.*

Remark 2.

1. It is known that for given $\epsilon > 0$ there exist finitely many ϵ-brackets $[l_k, u_k]$ whose union contains $\mathcal{G} = \{f_t : t \in [0, 1]\}$ and such that $\|u_k - l_k\| < \epsilon$ for every $k = 1, \ldots, N_{[\,]}(\epsilon)$. This is possible because the cardinality of $\mathcal{G}$ is the same as that of $[0, 1]$. See Van der Vaart and Wellner (1996)

2. It is also known that for given $\epsilon > 0$ there exist finitely many ϵ-brackets $[l_k, u_k]$ whose union contains $\mathcal{G} \otimes [0, 1] = \{(f_t, s) : t \in [0, 1], s \in [0, 1]\}$ and such that $\|u_k - l_k\| < \epsilon$ for every $k = 1, \ldots, N_{[\,]}(\epsilon)$. This is possible because the cardinality of $\mathcal{G} \otimes [0, 1]$ is the same as that of $[0, 1] \otimes [0, 1]$.

Proof: (Proof of Theorem 1) We observe that the filtration M_i becomes

$$M_i := \sigma\Big((x_j, y_j) : j \le i\Big).$$

We also observe that for each $t \in [0,1]$,

$$E(f_t(x_i, y_i)|\, M_{i-1}) = E(f_t(x_0, y_0)|\, M_{-1}) = E(f_t(x,y)\,|y) = 0$$

as follow from

$$E(f_t(x,y)|\, y) = 1_{[0,t]}(y)E1_{[0,\frac{1}{2})}(x) - 1_{[0,t]}(y)E1_{[\frac{1}{2},1)}(x) = 0.$$

This means that $\mathcal{G} = \{f_t(x,y) : t \in [0,1]\} \subseteq H_0 \ominus H_{-1}$. The fact that the baker transformation is ergodic appears in p.296 in Durrett (1991). Notice that the Lebesgue measure P is invariant under the baker transformation ϕ, *i.e.*, $P\phi^{-1} = P$. For each fixed $t \in [0,1]$, we apply the ergodic theorem to the baker transformation ϕ, see Durrett (1991), to get

$$B_n(t) = \frac{1}{n}\sum_{i=0}^{n-1} f_t(x_i, y_i) = \frac{1}{n}\sum_{i=0}^{n-1} f_t\left(\phi^i(x,y)\right) \to E\left(f_t(x,y)|\, y\right) = 0.$$

almost surely and in the mean. □

Proof: (Proof of Theorem 2.) Fix $\epsilon > 0$. Choose finitely many ϵ-brackets $[l_k, u_k]$ whose union contains $\mathcal{G}$ and such that $\|u_k - l_k\| < \epsilon$ for every $k = 1, \dots, N_{[\,]}(\epsilon)$. Then, for every $f_t \in \mathcal{G}$, there is a bracket such that

$$\begin{aligned} B_n(f_t) &= \frac{1}{n}\sum_{i=0}^{n-1} f_t(x_i, y_i) \\ &= \frac{1}{n}\sum_{i=0}^{n-1} f_t(x_i, y_i) - Eu_k(x,y) + Eu_k(x,y) \\ &\le \frac{1}{n}\sum_{i=0}^{n-1} f_t(x_i, y_i) + E|u_k - l_k|(x,y) \\ &\le \frac{1}{n}\sum_{i=0}^{n-1} u_k(x_i, y_i) + E|u_k - l_k|(x,y). \end{aligned}$$

Consequently,

$$\sup_{t\in[0,1]} B_n(f_t) \le \max_{1\le k\le N_{[\,]}(\epsilon)} \frac{1}{n}\sum_{i=0}^{n-1} u_k(x_i, y_i) + \epsilon.$$

The right hand side converges almost surely and in the mean to ϵ by Theorem 1. Combination with a similar argument for $\inf_{t\in[0,1]} B_n(f_t)$ yields that

$$\limsup_{n\to\infty} \sup_{t\in[0,1]} |B_n(f_t)|^* \le \epsilon,$$

almost surely, for every $\epsilon > 0$. Take a sequence $\epsilon_m \downarrow 0$ to see that the limsup must actually be zero almost surely. The proof Theorem 2 is completed.

□

Proof: (Proof of Theorem 3) The proof is exactly same as that of Theorem 2 by choosing the brackets as in the Remark 2. □

References

DeHardt, J. (1971). Generalizations of the Glivenko-Cantelli theorem, *The Annals of Mathematical and Statistics*, **42**, 2050–2055.

Durrett, R. (1991). *Probability: Theory and Examples*, Wadsworth, California.

Hoffmann-Jϕrgensen, J. (1991). Stochastic processes on Polish spaces, Aarhus Universitet, Matematisk Institut.

Van de Geer, S. (2000). *Empirical Processes in M-Estimation*, Cambridge University Press, Cambridge.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer-Verlag, New York.

Received September 2008; Accepted November 2008