ISRAEL JOURNAL OF MATHEMATICS 94 (1996), 179–188

AN EXAMPLE OF POINTWISE NON-CONVERGENCE OF ITERATED CONDITIONAL EXPECTATION OPERATORS

BY

MUSTAFA AKCOGLU

University of Toronto, Toronto, Canada
e-mail: akcoglu@math.toronto.edu

AND

JONATHAN L. KING*

University of Florida, Gainesville, FL 32611-2082, USA
e-mail: squash@math.ufl.edu
Web address: http://www.math.ufl.edu/~squash/

ABSTRACT

Given ε, we construct a sequence $\mathcal{F}_1, \mathcal{F}_2, \ldots$ of Borel sub-sigma-algebras on the unit interval with the following property. Suppose the identity function $f(x) = x$ is transformed by successive conditioning on $\mathcal{F}_1$, then $\mathcal{F}_2, \ldots$, then $\mathcal{F}_n, \ldots$. Then the lim sup, with respect to n, will exceed (pointwise almost-everywhere) $1-\varepsilon$ and its lim inf will be less than ε.

The sequence of functions also will fail to converge in the $\mathbb{L}_2$-norm. This contrasts with the long-open conjecture that if all the $\mathcal{F}_n$ come from a finite set of sigma-algebras, then the resulting sequence of functions must converge in $\mathbb{L}_2$.

1. Preface

In 1961 D. L. Burkholder and Y. S. Chow proved a basic result [4] that started a very interesting line of research which is still continuing. They showed that if E and F are two conditional expectations on a probability space and f is an $\mathbb{L}_2$-function, then $(EF)^n f$ converges pointwise (a.e.). A related but different theorem of E. Stein [7] (see also [8]) implies this result for $1 < p < \infty$ (and proves

* J. L. King was partially supported by NSF grant DMS-9112595.
Received July 28, 1994

more in $\mathbb{L}_2$). In 1962 G.-C. Rota [6] obtained a very general and elegant result, the "Alternierende Verfahren" theorem, that contained the result of Burkholder and Chow. See [1] for further results in this direction. On the other hand, D.S. Ornstein gave an example [5] to show that $(EF)^n f$ may diverge for some $\mathbb{L}_1$-functions.

A natural problem is to consider more than two conditional expectations. Rota's theorem gives the convergence for certain types of sequences obtained in terms of a sequence of conditional expectations. Recently A. Brunel [3] announced a deep new result: if $f \in \mathbb{L}_2$ then $(E_1 E_2 \dots E_k)^n f$ converges pointwise (convergence in the $\mathbb{L}_2$-norm was already known), where E_i's are finitely many conditional expectations. His methods appear to be very intricate, and in the general spirit of Stein's methods in [7]. The general problem of convergence under the application of a sequence of conditional expectations seems to be still open, even for norm convergence in $\mathbb{L}_2$. Under the assumption that only finitely many conditional expectations are used, however, I. Amemiya and T. Ando [2] were able to obtain weak convergence in $\mathbb{L}_2$. In this note we will show that this finiteness assumption cannot be relaxed for any of these types of convergence by constructing a sequence of conditional expectations $\{E_n\}_1^\infty$ and a bounded function f such that $E_n \dots E_1 f$ diverges pointwise.

OVERVIEW. Suppose $X \subset \mathbb{R}$ has finite measure, μ is Lebesgue measure on X, and $\mathcal{X}$ is the Lebesgue field (sigma-algebra) on X. Given a subfield $\mathcal{F} \subset \mathcal{X}$, let $E_\mathcal{F}$ denote the conditional expectation operator.

Suppose W is a finite or infinite sequence of fields

$$W = (\dots, \mathcal{F}_n, \dots, \mathcal{F}_3, \mathcal{F}_2, \mathcal{F}_1), \qquad \text{for } 1 \le n < M.$$

For each n less than M, let $W_{(n)}$ denote the finite sequence $(\mathcal{F}_n, \dots, \mathcal{F}_1)$. Let $W_{(n)}$ also denote the iterated conditional-expectation operator

$$W_{(n)} := E_{\mathcal{F}_n} E_{\mathcal{F}_{n-1}} \cdots E_{\mathcal{F}_2} E_{\mathcal{F}_1},$$

which acts on $\mathbb{L}_1(X)$.

A finite sequence, W, of fields will be called a **word**; an infinite sequence is an **infinite-word**. The letters U, V, W will be used for words.

In order to state the theorem, we define two functions. The "orbit supremum" function is, pointwise, the supremum

$$\text{orb-Sup}\, Wf := \sup_{n:\, n<M} W_{(n)} f.$$

Define "orb-Inf" analogously. If W is an infinite-word, allow the abbreviation

$$\text{orb-Limsup}\, Wf \;:=\; \limsup_{n\to\infty} W_{(n)}f$$

and similarly for "orb-Liminf".

The symbols ε, δ and α always denote numbers in $\left(0, \frac{1}{2}\right)$.

THEOREM: *Given any positive α, there exists an infinite-word W such that for the identity function $f(x) = x$ on the unit interval $\left[-\frac{1}{2}, +\frac{1}{2}\right)$,*

$$\text{orb-Limsup}\, Wf \geq +\left(\tfrac{1}{2} - \alpha\right)$$

and

$$\text{orb-Liminf}\, Wf \leq -\left(\tfrac{1}{2} - \alpha\right)$$

almost-everywhere.

CONVENTIONS. The symbol "$\bigsqcup$" means "disjoint union". Unless specified otherwise, any "interval" is left-closed and right-open. The only functions we consider are especially simple piecewise linear functions; functions of the form $h\colon X \to \mathbb{R}$, where $X = \bigsqcup_k I_k$ is a finite disjoint union of intervals and where each $h|_{I_k}$ has constant slope 1.

When referring to a subset G of the domain of h, let the complement "G^c" mean $\mathrm{Domain}(h) \setminus G$.

Fields. Whenever we define a field, $\mathcal{F}$, on a subset G of $\mathbb{R}$, agree to extend $\mathcal{F}$ to be a field on $\mathbb{R}$ by assuming that every measurable subset of $\mathbb{R} \setminus G$ is in $\mathcal{F}$. Thus, if we have a function h on X, and on some subset $G \subset X$ we specify a field $\mathcal{F}$, then "$E_{\mathcal{F}}h$" makes sense and

$$E_{\mathcal{F}}h(x) \;=\; h(x)\,, \quad \text{for any } x \in G^c.$$

Uniform distribution. A function $f\colon X \to \mathbb{R}$ is "uniform onto $[A, B)$" if, for any subinterval $I \subset [A, B)$, the measure of $f^{-1}(I)$ equals the length of I.

2. Establishing the theorem

The essential ingredient, a nested "divide and conquer" argument, is Lemma 2, which is proved in this section. Proofs of the other lemmata are deferred to Section 3.

In the following lemma, the domain of f is partitioned into left and right halves and then the distribution of f on the two halves is interchanged.

REVERSAL LEMMA, 1: *Suppose $f\colon L \sqcup R \to \mathbb{R}$, where L and R are disjoint equal-mass intervals. Suppose $f|_L$ is uniform onto $[-S-\delta, -\delta)$ and $f|_R$ is uniform onto $[\delta, S+\delta)$. Suppose that 2δ, the "gap" between the uniform distributions, divides the length S.*

Then there is a word V such that $(Vf)|_L$ is uniform onto $[-\delta, S-\delta)$ and $(Vf)|_R$ is uniform onto $[-S+\delta, +\delta)$.

In Figure 3, the transition from panel **(b)** to panel **(c)** illustrates the Reversal Lemma. Our next lemma refers to the following property:

> STATEMENT $P(\varepsilon)$: *Suppose f is uniform onto interval $[A, B)$, whose length is $S := B - A$.*
>
> *Then, for any small positive δ (called the "shrinkage") there exists a word U and a set $G \subset \text{Domain}(f)$ so that $\mu(G^c) = 2\delta$ and for each $x \in G$:*
>
> $$\big[\text{orb-Sup}\, Uf\big](x) \geq B - \varepsilon S$$
>
> *and*
>
> $$\big[\text{orb-Inf}\, Uf\big](x) \leq A + \varepsilon S.$$
>
> *Moreover, on G our new function Uf is distributed uniformly onto the interval $[A+\delta, B-\delta)$.*

LEMMA 2: *$P(\varepsilon)$ holds for all positive ε.*

Evidently $P(1)$ holds. Thus this lemma will follow from the following implication.

Proof that $P(\varepsilon) \implies P(\frac{2}{3}\varepsilon)$: Without loss of generality, the domain and range of f are both $\left[-\frac{1}{2}, \frac{1}{2}\right)$ and so $S = 1$.

Assuming that $P(\varepsilon)$ holds and given a desired shrinkage δ_0, we wish to demonstrate $P(\frac{2}{3}\varepsilon)$. Pick a positive $\delta \ll \delta_0$, which will be specified as the argument proceeds. Figure **3(a)** shows the uniformly distributed function $f(x) = x$ on the interval $\left[-\frac{1}{2}, \frac{1}{2}\right)$; let L and R denote the left and right halves of the interval. As the construction progresses clockwise through the four panels of Figure 3, sets L and R will shrink slightly in that mass 4δ will be removed from them; nonetheless, we will continue to call them L and R.

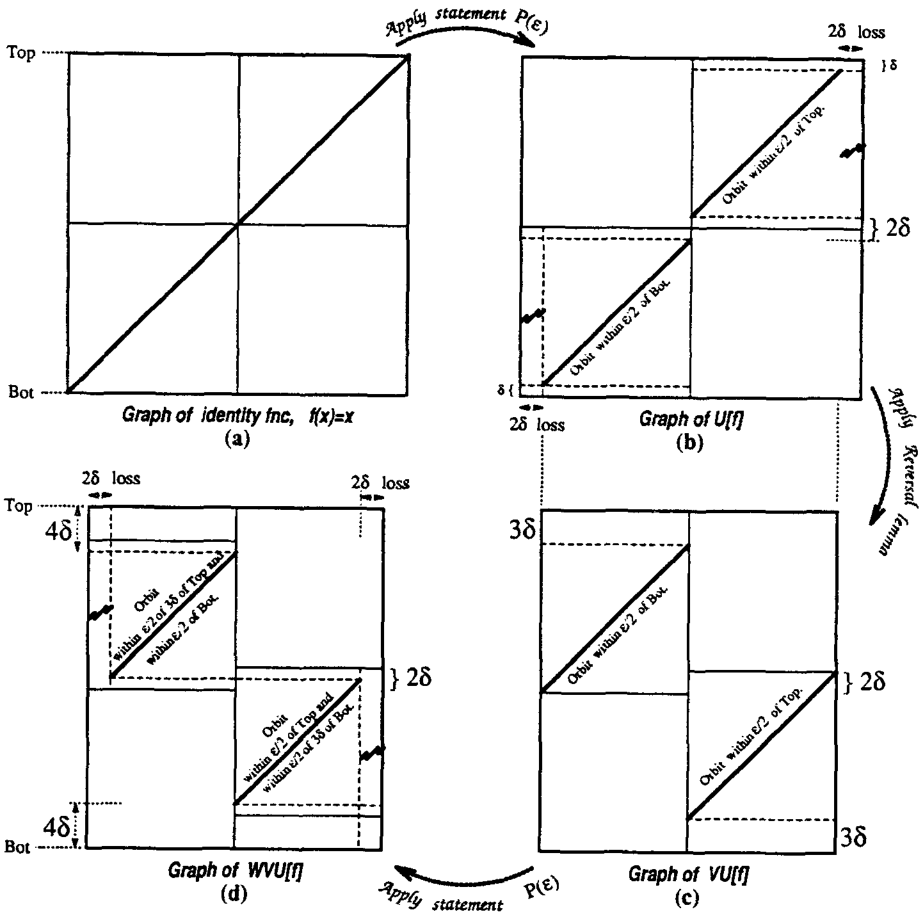

Figure 3

In the figure, we use Top $= +\frac{1}{2}$ and Bot $= -\frac{1}{2}$ to indicate the top and bottom of the range of f.

3(a) to **3(b)**: To L and R, apply $P(\varepsilon)$ with δ-shrinkage. This produces a word U so that

$$[\text{orb-Inf}\, Uf](x) \quad \text{is within } \varepsilon/2 \text{ of Bot}$$

(since $\varepsilon \cdot [\frac{1}{2}S] = \varepsilon/2$), for all $x \in L$ except for a "discarded" set of mass 2δ. This discarded set, which is actually a finite union of subintervals scattered throughout L, is drawn in Figure **3(b)** at the left side of L.

There is a similar romanticization in the graph of Uf over L; it is indeed uniformly distributed, but is not actually the identity function as shown.

3(b) to **3(c)**: The "discarded" parts of L and R are removed from the picture. Then we apply a word V to re<u>v</u>erse the left and right halves of the graph; the left half moves up and the right half drops.

3(c) to **3(d)**: To the left and right halves of **3(c)**, apply $P(\varepsilon)$ with δ-shrinkage. This produces a word, call it W, which shrinks the mass of L by another 2δ and arranges that for every x in this smaller L,

$$\big[\text{orb-Sup}\, WVUf\big](x) \quad \text{is within } (\varepsilon/2) + 3\delta \text{ of Top.}$$

Since δ can have been chosen so that $\frac{\varepsilon}{2} + 3\delta < \frac{2}{3}\varepsilon$, we have now arranged that, for $x \in L$,

$$\big[\text{orb-Sup}\, WVUf\big](x) \geq \text{Top} - \tfrac{2}{3}\varepsilon$$

and

$$\big[\text{orb-Inf}\, WVUf\big](x) \leq \text{Bot} + \tfrac{2}{3}\varepsilon.$$

The same holds for x in R, the (slightly shrunken) right half.

The final step: The function $h := WVUf$, when restricted to $L \sqcup R$, is necessarily uniform onto $\big[\text{Bot} + 4\delta, \text{Top} - 4\delta\big)$. Without loss of generality, $4\delta < \delta_0$. Thus, if we define G to be the set of $x \in L \sqcup R$ such that

$$h(x) \;\in\; \big[\text{Bot} + \delta_0, \text{Top} - \delta_0\big),$$

then $h|_G$ is indeed uniformly distributed onto $\big[\text{Bot} + \delta_0, \text{Top} - \delta_0\big)$, as desired. ∎

We now need to embellish the above lemma so that it applies to functions which are not quite uniformly distributed.

Definition: Say that $f\colon X \to \mathbb{R}$ is "nicely distributed on $[A, B)$" if there is a subset $G \subset X$ such that

- $f|_G$ is uniform onto $[A, B)$.
- The complement, $X \setminus G$, is a finite disjoint union of intervals, $\bigsqcup_1^K I_k$, such that each $f|_{I_k}$ is uniform onto some subinterval of the open interval (A, B).

LEMMA 4: *Suppose $f\colon X \to \mathbb{R}$ is nicely distributed onto $[A, B)$. Then for any small ε, there is a word U so that Uf is nicely distributed on $[A+\varepsilon, B-\varepsilon)$. Moreover,*

$$\begin{aligned}&\big[\text{orb-Sup}\, Uf\big](x) \geq B - \varepsilon \\ (5) \qquad \textit{and} \qquad & \\ &\big[\text{orb-Inf}\, Uf\big](x) \leq A + \varepsilon,\end{aligned}$$

for all $x \in X$ off a set of mass less than ε.

The benefit of this lemma is that ε can be chosen arbitrarily small compared to the measure of the intervals I_k. Iteratively using this lemma establishes the theorem, as follows.

Proof of the Theorem: Set $A_0 := \frac{1}{2}$ and $f_0 := f$.

For $n = 1, 2, \ldots$, let $\varepsilon_n := \alpha/2^n$. Since f_{n-1} is nicely distributed onto $\big[-A_{n-1}, A_{n-1}\big)$, we can apply Lemma 4 to f_{n-1}, with $\varepsilon = \varepsilon_n$. This produces a function

$$f_n := U_n f_{n-1}$$

which is nicely distributed onto $[-A_n, A_n)$, where $A_n := A_{n-1} - \varepsilon_n$. Moreover, since $\varepsilon_1 + \cdots + \varepsilon_n$ is less than α,

$$\begin{aligned}&\big[\text{orb-Sup}\, U_n f_{n-1}\big](x) \;\geq\; +\frac{1}{2} - \alpha \\ (6[n]) \qquad \text{and} \qquad & \\ &\big[\text{orb-Inf}\, U_n f_{n-1}\big](x) \;\leq\; -\frac{1}{2} + \alpha\end{aligned}$$

for all x off a "bad set" whose mass is less than ε_n.

Since the masses of the bad sets are summable, the Borel-Cantelli lemma tells us –after having deleted a nullset– that for every x, inequality (6[n]) holds for infinitely many n.

Letting W be the infinite-word $\cdots U_3 U_2 U_1$, then, gives us the conclusion of the theorem.

3. Details

We now sketch the minutia of the two unproved lemmata of the preceding section.

Notice that if the procedure used to prove $P(\varepsilon) \implies P(\frac{2}{3}\varepsilon)$ is preceded by "Reversal", that is, it becomes $[\text{Reversal}, P(\varepsilon), \text{Reversal}, P(\varepsilon)]$, then the resulting function Uf will be nicely distributed.

Thus we may henceforth assume, when we "apply $P(\alpha)$" to a function f, that the output function Uf is nicely distributed, and that shrinkage $= \alpha$ was used.

Sketch of proof of Lemma 4: Recall the set G and intervals $\bigsqcup_1^K I_k$ from the definition of "nicely distributed".

Suppose we take a number α which is infinitesimal compared to the given ε. Apply $P(\alpha)$ to $f|_G$. This will produce an infinitesimally smaller set G and infinitesimally smaller interval $[A, B)$, so that (5) holds for all $x \in G$ except for a set of infinitesimal mass (mass $= 2\alpha$). The upshot is, we really only have to establish (5) for most points $x \in \bigsqcup_1^K I_k$.

Since the number of intervals, K, is known in advance, it suffices to take a single interval $I = I_k$ and show how to establish (5) for all $x \in I$ except for a set of infinitesimal mass.

Again, fix a number $\alpha \ll \varepsilon$. Since $f|_I$ is uniform onto some subinterval of (A, B), there is a subset $J \subset G$ on which f has the same distribution as on I. Thus

$$f|_{(G \setminus J) \sqcup I} \quad \text{is uniform onto } [A, B).$$

Now apply $P(\alpha)$ to this $f|_{(G \setminus J) \sqcup I}$. For all points $x \in I$ except for a set of infinitesimal mass, inequality (5) holds. And this operation only infinitesimally changes G and $[A, B)$. ∎

Proof of the Reversal Lemma: There is no loss of generality in assuming that the gap, 2δ, equals 2. Nor in assuming that

$$L = \bigsqcup_{k=-K}^{-1} I_k \qquad \text{and} \qquad R = \bigsqcup_{k=1}^{K} I_k,$$

where I_k is the interval $[2k-1, 2k+1)$.

Given distinct indices k and ℓ, let $\mathcal{F}_{k,\ell}$ be the field which links corresponding points in I_k and I_ℓ. That is, a subset $D \subset I_k \sqcup I_\ell$ is in $\mathcal{F}_{k,\ell}$ precisely when, for every x,

$$2k + x \in D \iff 2\ell + x \in D.$$

On the complement of $I_k \sqcup I_\ell$, as per our convention, $\mathcal{F}_{k,\ell}$ agrees with the full Lebesgue field on $\mathbb{R}$.

Now suppose $f|_{I_k}$ and $f|_{I_\ell}$ are slope 1 onto intervals of the form $\big[A-3, A-1\big)$ and $\big[A+1, A+3\big)$, respectively. Condition f on $\mathcal{F}_{k,\ell}$ and call the resulting function h. Then

$$h|_{I_k} \text{ and } h|_{I_\ell} \text{ are both slope 1 onto } \big[A-1, A+1\big).$$

Thus the effect of conditioning is to raise the distribution on I_k by the gap length, and to lower the distribution on I_ℓ by the gap length.

Consequently, one can prove the Reversal Lemma by starting with $f(x) = x$ and conditioning successively on $\mathcal{F}_{k,\ell}$ as the pair (k, ℓ) takes on these values:

$$\begin{aligned}
&(-1,1),\ (-1,2),\ (-1,3),\ \ldots,\ (-1,K),\\
&\text{and continue,}\\
&(-2,1),\ (-2,2),\ (-2,3),\ \ldots,\ (-2,K),\\
&(-3,1),\ (-3,2),\ (-3,3),\ \ldots,\ (-3,K),\\
&\quad\vdots\\
&(-K,1),\ (-K,1),\ (-K,1),\ \ldots,\ (-K,K).
\end{aligned}$$

The successive conditioning on these K^2 many fields essentially exchanges the distributions on L and R. ∎

References

[1] M. A. Akcoglu, J. R. Baxter and W. F. M. Lee, *Representation of positive operators and alternating sequences*, Advances in Mathematics **87** (1991), 249–290.

[2] I. Amemiya and T. Ando, *Convergence of random products of contractions in Hilbert Space*, Acta Scientiarum Mathematicarum (Szegeed) **26** (1965), 239–244. Mathematical Review: The review has interesting mathematics. (See also Mathematical Reviews, MR v.32, #4570.)

[3] A. Brunel, Private communication.

[4] D. L. Burkholder and Y. S. Chow, *Iterates of conditional expectations*, Proceedings of the American Mathematical Society **12** (1961), 490–495.

[5] D. S. Ornstein, *On the pointwise behaviour of iterates of a self-adjoint operator*, Journal of Mathematical Mechanics **18** (1968), 473–478.

[6] G.-C. Rota, *An "Alternierende Verfahren" for general operators*, Bulletin of the American Mathematical Society **68** (1962), 95–102.

[7] E. Stein, *On the maximal ergodic theorem*, Proceedings of the National Academy of Sciences of the United States of America **47** (1961), 1894–1897.

[8] E. Stein, *Topics in Harmonic Analysis*, Princeton University Press, 1970.