A NONSTANDARD PROOF OF THE MARTINGALE CONVERGENCE THEOREM

LESTER L. HELMS* AND PETER A. LOEB*

In this note we use A. Robinson's [5] nonstandard analysis to give an elementary proof of the almost everywhere convergence of an L^1 -bounded submartingale. Here, the index set \mathscr{I} is a countable subset of the real numbers **R**; we assume that \mathscr{I} contains the natural numbers **N**, but any cofinal subset of **R** will do. The continuous parameter martingale convergence theorem usually reduces to the case considered here. Our proof does not use the notion of a stopping time. It does employ a nonstandard criterion for almost everywhere convergence and demonstrates the usefulness of that criterion. It also produces the limit function.

We shall use the notation from [4] to which we refer the reader for further details about nonstandard analysis in general. We assume that we are working with a fixed \aleph_1 -saturated, nonstandard extension of a standard structure. Of course, ***R** and ***N** denote the nonstandard extensions of **R** and **N**, and $a \cong b$ means that a - b is infinitesimal in ***R**. If (X, \mathscr{F}, μ) is an internal measure space and $g: X \to *\mathbf{R} \cup \{-\infty, +\infty\}$ is internal and \mathscr{F} -measurable, then (following K. Stroyan) we shall say that $g \cong 0$ nearly surely (n.s.) when the following holds: For some infinitesimal $\varepsilon > 0$, $\mu(|g| > \varepsilon) \cong 0$. Clearly, $g \cong 0$ n.s. if and only if for each $\varepsilon > 0$ in **R**, $\mu(|g| > \varepsilon) < \varepsilon$.

We now establish a nonstandard criterion for almost everywhere convergence. Here, as later, \mathscr{I} denotes a countable subset of \mathbf{R} with $\mathbf{N} \subset \mathscr{I}$. The ordering on \mathscr{I} is the ordering inherited from \mathbf{R} . We shall use n, m, and k to denote natural numbers, while i and j will denote elements of \mathscr{I} or $*\mathscr{I}$. Moreover, $\{i: n \leq i \leq m\}$ will denote the set of indices in just \mathscr{I} with $n \leq i \leq m$, while if γ and η are in $*\mathbf{N} - \mathbf{N}$, then $\{i: \gamma \leq i \leq \eta\}$ will denote the set of indices in $*\mathscr{I}$ with $\gamma \leq i \leq \eta$. Given $n \in \mathbf{N}$, $\bigcup_{i \geq n} A_i$ will denote $\bigcup \{A_i: i \in \mathscr{I}, i \geq n\}$.

THEOREM 1. Let (X, \mathcal{F}, μ) be a standard measure space with $\mu(X) < +\infty$, and for each $i \in \mathcal{I}$, let g_i be an extended real-valued, \mathcal{F} -measurable function on X.

^{*}This research was supported in part by a grant from the U.S. National Science Foundation (NSF MCS 76–07471).

Received by the editors on July 11, 1980, and in revised form on October 14, 1980.

Copyright © 1982 Rocky Mountain Mathematics Consortium

i) If $g_i \to 0$ a.e., then for each γ and η in *N - N with $\gamma \leq \eta$,

$$\sup_{\gamma \leq i \leq \eta} |g_i| \cong 0 \text{ n.s.}$$

ii) Conversely, $g_i \rightarrow 0$ a.e. if there is an $\eta \in *N - N$ such that for all $\gamma \leq \eta$ in *N - N,

$$\sup_{\gamma \leq i \leq \eta} |g_i| \cong 0 \text{ n.s.}$$

iii) Assume $\mathcal{I} = \mathbf{N}$. Then there is a null set $A \subset X$ such that $g_i(x)$ is a Cauchy sequence in \mathbf{R} for each $x \in X - A$ if and only if the following holds: For some $\eta \in *\mathbf{N} - \mathbf{N}$ and all $\gamma \leq \eta$ in $*\mathbf{N} - \mathbf{N}$,

$$\max_{\substack{\gamma \leq i \leq j \leq \eta}} |g_i - g_j| \cong 0 \text{ n.s.}$$

PROOF. (i) If $g_i \to 0$ a.e., then given $\varepsilon > 0$ in **R** and setting $A_i^{\varepsilon} = \{|g_i| > \varepsilon\}$ for each $i \in \mathcal{I}$, we have

$$\mu \left(\bigcap_{k=1}^{\infty} \bigcup_{i\geq k} A_i^{\varepsilon} \right) = 0,$$

whence $\lim_{k\to\infty} \mu(\bigcup_{i\geq k} A_i^{\varepsilon}) = 0$. It follows that for $\gamma \leq \eta$ in *N – N, * $\mu(\bigcup_{\gamma\leq i\leq \eta} A_i^{\varepsilon}) \cong 0$, and so * $\mu(\sup_{\gamma\leq i\leq \eta} |g_i| > \varepsilon) < \varepsilon$. Since this is true for any $\varepsilon > 0$ in **R**, $\sup_{\gamma\leq i\leq \eta} |g_i| \cong 0$ n.s.

(ii) If there is an $\eta \in {}^{*}\mathbf{N} - \mathbf{N}$ such that for each $\gamma \leq \eta$ in ${}^{*}\mathbf{N} - \mathbf{N}$ and each $\varepsilon > 0$ in \mathbf{R} we have ${}^{*}\mu(\sup_{\gamma \leq i \leq \eta} |g_i| > \varepsilon) \cong 0$ whence ${}^{*}\mu(\bigcup_{\gamma \leq i \leq \eta} A_i^{\varepsilon})$ $\cong 0$, then it follows that for any $\delta > 0$ in \mathbf{R} there is a $k \in \mathbf{N}$ with $\mu(\bigcup_{i \geq k} A_i^{\varepsilon})$ $\leq \delta$. Therefore, $\mu(\bigcap_{k=1}^{\infty} \bigcup_{i \geq k} A_i^{\varepsilon}) = 0$. Let

$$B = \bigcup_{m=1}^{\infty} \bigcap_{k=1}^{\infty} \bigcup_{i \ge k} A_i^{1/m}.$$

Then $\mu(B) = 0$ and $g_i(x) \to 0$ for $x \in X - B$.

(iii) The proof is left to the reader; we shall not need this result.

The above criterion for almost everywhere convergence was suggested by Egorov's Theorem. A sufficient condition based on the Borel-Cantelli Lemma was used by Hersh and Greenwood [3] to consider the convergence of L^2 -bounded martingales. A "maximal" condition similar to the one used here does appear in the body of their proof; further use of this maximal condition seems to be needed to carry out their proof.

We now fix an increasing family of σ -algebras $\{\mathscr{F}_i: i \in \mathscr{I}\}$ in a standard set X. We let \mathscr{F} denote the smallest σ -algebra containing each \mathscr{F}_i , i.e., $\mathscr{F} = \sigma(\bigcup_{i \ge 1} \mathscr{F}_i) = \sigma(\bigcup_{n=1}^{\infty} \mathscr{F}_n)$. Let P be a fixed probability measure on (X, \mathscr{F}) . Let \mathscr{I} be a finite or infinite subset of \mathscr{I} . Recall that a family $\{Y_j: j \in \mathscr{J}\}$ of extended real-valued functions on X is said to be adapted to $\{\mathscr{F}_j: j \in \mathscr{J}\}$ if Y_j is \mathscr{F}_j -measurable for each $j \in \mathscr{J}$. If, moreover,

166

 $\int_A Y_{i_0} dP \leq \int_A Y_{j_0} dP$ (or, respectively, $\int_A Y_{i_0} dP = \int_A Y_{j_0} dP$) for each pair $i_0 < j_0$ in \mathscr{J} and each $A \in \mathscr{F}_{i_0}$, then $\{Y_j: j \in \mathscr{J}\}$ is called a submartingale (or, respectively, a martingale). For completeness, we prove the following inequalities of Doob.

THEOREM 2 (DOOB [2, p. 314]). Given n < m in N, let $\{Y_i: n \leq i \leq m\}$ be a submartingale adapted to $\{\mathcal{F}_i: n \leq i \leq m\}$. Fix $\lambda \in \mathbb{R}$ and $A \in \mathcal{F}_n$. Let $\overline{M} = \{\sup_{n \leq i \leq m} Y_i > \lambda\} - A$ and let $\underline{M} = \{\inf_{n \leq i \leq m} Y_i < \lambda\} - A$. Then

(i)
$$\lambda P(\overline{M}) \leq \int_{\overline{M}} Y_m dP$$
,
(ii) $\int_{X-A} Y_n dP - \lambda P(\underline{M}) \leq \int_{(X-A)-\underline{M}} Y_m dP$.

PROOF. We assume that $\mathscr{I} = \mathbf{N}$. The general case follows by taking appropriate limits with respect to increasing finite subsets of \mathscr{I} ; the order on the family of finite subsets of \mathscr{I} is given by containment.

(i) Define B_i by induction so that $B_n = \{Y_n > \lambda\} - A$ and for $n < i \le m$, $B_i = [\{Y_i > \lambda\} - \bigcup_{j=n}^{i-1} B_j] - A$. Then, using the submartingale property on $B_i \in \mathcal{F}_i$, we have

$$\int_{\overline{M}} Y_m dP = \int_{\substack{\bigcup \\ i = n} B_i} Y_m dP \ge \sum_{i = n}^m \int_{B_i} Y_i dP$$
$$\ge \sum_{i = n}^m \lambda P(B_i) = \lambda P(\overline{M}).$$

(ii) Define C_i and D_i by induction so that $C_n = \{Y_n < \lambda\} - A$, $D_n = \{Y_n \ge \lambda\} - A$, and for $n < i \le m$, $C_i = \{x \in D_{i-1}: Y_i < \lambda\}$ and $D_i = D_{i-1} - C_i$. Then

$$\int_{X-A} Y_n dP \leq \lambda P(C_n) + \int_{D_n} Y_n dP$$

$$\leq \lambda P(C_n) + \int_{D_n} Y_{n+1} dP$$

$$\leq \lambda P(C_n) + \lambda P(C_{n+1}) + \int_{D_{n+1}} Y_{n+1} dP$$

$$\vdots$$

$$\leq \lambda P(\underline{M}) + \int_{D_m} Y_m dP.$$

Recall that for $Y \in L^1(X, \mathcal{F}, P)$, $E[Y|\mathcal{F}_i] \in L^1(X, \mathcal{F}_i, P)$ is the Radon-Nikodym derivative of the measure obtained by integrating Y over sets in \mathcal{F}_i . The following result is proved here using a simplification of a proof by Alda [1].

THEOREM 3. Let Y be \mathcal{F} -measurable and integrable on X. Then $E[Y|\mathcal{F}_i] \rightarrow Y$ a.e.

PROOF. Fix γ and η in *N - N with $\gamma < \eta$. Since $\mathscr{F} = \sigma(\bigcup_{n=1}^{\infty} \mathscr{F}_n)$, there is for any standard set $A \in \mathscr{F}$ and any $\varepsilon > 0$ an $n \in N$ and an $E \in \mathscr{F}_n$ such that $P(A \triangle E) < \varepsilon$, and so there is an $E_A \in \mathscr{F}_{\gamma}$ with $*P(*A \triangle E_A) \cong 0$. It follows that there is an internal \mathscr{F}_{γ} -measurable function Y_0 on *X such that $\int_{*X} |*Y - Y_0| d^*P \cong 0$, whence $|*Y - Y_0| \cong$ 0 n.s. Let $Y_i = E[*Y|\mathscr{F}_i]$ for $\gamma \leq i \leq \eta$. Then $\{|Y_i - Y_0|: \gamma \leq i \leq \eta\}$ is an internal submartingale and this is also true if $|*Y - Y_0|$ is adjoined as a last element. Given $\varepsilon > 0$ in **R** and setting $B_{\varepsilon} = \{\sup_{\gamma \leq i \leq \eta} | Y_i - Y_0| > \varepsilon\}$, we have by (i) of Theorem 2,

$$\varepsilon^* P(B_{\varepsilon}) \leq \int_{B_{\varepsilon}} |Y_{\eta} - Y_0| \ d^* P \leq \int_{B_{\varepsilon}} |*Y - Y_0| \ d^* P \simeq 0.$$

Therefore, $P(B_{\varepsilon}) \cong 0$ for any $\varepsilon > 0$ in **R**, and so $\sup_{\gamma \le i \le \eta} |Y_i - Y_0| \cong 0$ n.s. Since $|Y_0 - *Y| \cong 0$ n.s., $\sup_{\gamma \le i \le \eta} |Y_i - *Y| \cong 0$ n.s., whence $|Y_i - Y| \to 0$ a.e.

We now consider a fixed $\{\mathscr{F}_i\}$ -adapted submartingale $\{Y_i: i \in \mathscr{I}\}$ such that $Y_i \ge 0$ for each $i \in \mathscr{I}$ and $L = \sup_i \int_X Y_i dP < +\infty$. Since $\int_X Y_i dP \le \int_X Y_j dP$ when i < j in \mathscr{I} , $L = \lim_{i \to \infty} \int_X Y_i dP$. Thus $L = \circ \int_{*X} Y_i d^*P$ for each infinite $i \in *\mathscr{I}$. Given any infinite $i \in *\mathscr{I}$, we let

$$S_i = L - \lim_{\substack{m \to \infty \\ m \in N}} \int_{*X} (Y_i \wedge m) \, d^* P.$$

We call S_i the singular part of the integral of Y_i . One can find an $\alpha \in *\mathbf{N} - \mathbf{N}$ such that $S_i = \circ \int_{\langle Y_i \ge \alpha \rangle} Y_i d^*P$.

PROPOSITION 1. There is an $\eta \in *\mathbf{N} - \mathbf{N}$ such that for each infinite $i \leq \eta$ in $*\mathcal{I}$, $S_i = S_{\eta}$.

PROOF. Given an infinite $i \in \mathscr{I}$, choose $\alpha \in \mathsf{N} - \mathbb{N}$ so that $S_i = \circ \int_{\langle Y_i \geq \alpha \rangle} Y_i d^*P$. The set $D_i = \{Y_i \geq \alpha\}$ has infinitesimal measure, so for $j \geq i$ in \mathscr{I} and $m \in \mathbb{N}$, $\circ \int_{D_i} (Y_j \wedge m) d^*P = 0$, whence

$$S_i = \int_{D_i}^{\circ} Y_i \, d^* P \leq \int_{D_i}^{\circ} Y_j \, d^* P \leq S_j.$$

Let $\{\gamma_n: n \in \mathbb{N}\}$ be a decreasing sequence in $\mathbb{N} - \mathbb{N}$ such that $\lim_{n\to\infty} S_{\gamma_n} = \inf\{S_i: i \text{ infinite in } \mathscr{I}\}$. By \aleph_1 -saturation, there exists an $\eta \in \mathbb{N} - \mathbb{N}$ so that $\eta \leq \gamma_n$ for all $n \in \mathbb{N}$. Clearly $S_i = S_\eta$ for each infinite $i \leq \eta$ in $\mathbb{I}\mathcal{I}$.

We now fix $\eta \in {}^*\mathbb{N} - \mathbb{N}$ so that $S_i = S_{\eta}$ for any infinite $i \leq \eta$ in ${}^*\mathcal{I}$. Let $S = S_{\eta}$. If S = 0, we set $A_i = \emptyset$ for all $i \in {}^*\mathcal{I}$. If S > 0 and $i \leq \eta$ in ${}^*\mathcal{I}$, we let α_i be the largest element $\rho \in {}^*\mathbb{N}$ such that

$$\int_{\langle Y_i \ge \rho \rangle} Y_i \, d^* P \ge S - S/\rho,$$

and we let $A_i = \{Y_i \ge \alpha_i\}$. The proof of Proposition 1 shows that for each infinite $i \le \eta$ in * \mathscr{I} we have $\alpha_i \in *\mathbb{N} - \mathbb{N}, \int_{A_i} Y_i d^*P \cong \int_{A_i} Y_\eta d^*P \cong S, *P(A_i) \cong 0$, and $\int_{A_n-A_i} Y_\eta d^*P \cong 0$.

Given any set $B \in \mathscr{F}$, there is a set $E_B \in \mathscr{F}_{\eta}$ such that $*P(*B \triangle E_B) \cong 0$. Let $\nu(B) = \circ \int_{E_B - A_{\eta}} Y_{\eta} d*P$. Since

$$\lim_{n\to\infty} \int_{\langle Y_\eta\geq m\rangle-A_\eta}^{\circ} Y_\eta d^*P = 0,$$

there exists for each $\varepsilon > 0$ in **R** $a \ \delta > 0$ in **R** such that $\nu(B) < \varepsilon$ when $P(B) < \delta$. Thus ν is σ -additive on \mathscr{F} and absolutely continuous with respect to *P*. Let *Z* be the Radon-Nikodym derivative $d\nu/dP$.

PROPOSITION 2. The nonnegative submartingale $Y_i \rightarrow Z$ a.e.

PROOF. For each $i \in \mathscr{I}$, let $Z_i = E[Z|\mathscr{F}_i]$. By Theorem 3, $Z_i \to Z$ a.e.; we will show that $Y_i - Z_i \to 0$ a.e. Given $\varepsilon > 0$ in **R**, it follows from the properties of the A_i 's that there exists an $n \in \mathbb{N}$ such that $P(A_n)$ $< \varepsilon/2$ and $\circ \int_{A_\eta \to A_n} Y_\eta d^*P < \varepsilon^2/2$. For each $m \ge n$ in **N**, set $B_m =$ $\{\sup_{n \le i \le m} (Y_i - Z_i) > \varepsilon\} - A_n$. Since $Y_i - Z_i$ is a submartingale and $*B_m \in *\mathscr{F}_m \subset \mathscr{F}_n$, it follows from the definition of ν that

$$\varepsilon P(B_m) \leq \int_{B_m} (Y_m - Z_m) dP \leq \int_{*B_m} Y_\eta d*P - \int_{B_m} Z dP$$
$$\leq \int_{A_\eta - *A_n} Y_\eta d*P + \nu(B_m) - \int_{B_m} Z dP < \frac{\varepsilon^2}{2}.$$

Therefore, for any $m \ge n$ in N, $P(\sup_{n\le i\le m}(Y_i - Z_i) > \varepsilon) \le P(B_m) + P(A_n) < \varepsilon$. It follows that for $\gamma \le \eta$ in *N - N, $*P(\sup_{\gamma \le i\le \eta}(Y_i - Z_i) > \varepsilon) < \varepsilon$ for each $\varepsilon > 0$ in **R** and thus for some positive $\varepsilon \simeq 0$ in *R. On the other hand, if for each $\varepsilon > 0$ in **R** we set $M_{\varepsilon} = \{\inf_{\gamma \le i\le \eta}(Y_i - Z_i) < -\varepsilon\} - A_{\tau}$, then

$$\int_{*X-A_{\tau}} (Y_{\tau} - Z_{\tau}) d^*P + \varepsilon^* P(M_{\varepsilon}) \leq \int_{(*X-A_{\tau})-M_{\varepsilon}} (Y_{\eta} - Z_{\eta}) d^*P.$$

Here, $\int_{X-A_r} (Y_r - Z_r) d^*P \cong \nu(X) - \int_X Z dP = 0$. Moreover, we have just shown that $Y_\eta - Z_\eta$ is either negative or infinitesimal except on a set E of infinitesimal measure. Since $\int_{E-A_r} Y_\eta d^*P \cong 0$,

$$\int_{(*X-A_{\gamma}-M_{\ell})} (Y_{\eta}-Z_{\eta}) d^*P \leq 0.$$

Therefore, $*P(M_{\varepsilon}) \cong 0$ for each $\varepsilon > 0$ in **R**. It now follows that for any $\gamma \leq \eta$ in *N - N, $\sup_{\gamma \leq i \leq \eta} |Y_i - Z_i| \cong 0$ n.s., whence $Y_i - Z_i \to 0$ a.e.

It now follows from Proposition 2 that $S_i = S$ for every infinite $i \in *\mathscr{I}$. Moreover, when S = 0, the above proof shows that $P(\exists i \in \mathscr{I} \text{ with } Y_i > Z_i) = 0$. The nonnegative submartingale $\{Y_i\}$ is called uniformly integrable if S = 0 (see [4, Page 131]). We now prove the convergence theorem for L^1 -bounded submartingales.

THEOREM 4. Let $\{Y_i: i \in \mathcal{F}\}$ be an $\{\mathcal{F}_i\}$ -adapted submartingale with $\sup_i \int_X |Y_i| dP < +\infty$. Then Y_i converges a.e.

PROOF. For each $m \in \mathbb{N}$, $\{Y_i \lor -m\}$ is a submartingale as is $\{(Y_i \lor -m) + m\}$. By Proposition 2, $Y_i \lor -m$ converges a.e. for each $m \in \mathbb{N}$. Given $m \in \mathbb{N}$, let $D_m = \{\inf_{i \in \mathcal{J}} Y_i < -m\}$. For almost all $x \in X - \bigcap_{m=1}^{\infty} D_m$, $Y_i(x)$ converges. We will show that $P(\bigcap_{m=1}^{\infty} D_m) = 0$. Assume not. Then for some $\varepsilon > 0$ in \mathbb{R} , $P(D_m) \ge \varepsilon$ for all $m \in \mathbb{N}$ and thus for all $m \in \mathbb{N}$. Fix $m_0 \in \mathbb{N} - \mathbb{N}$. For each $\eta \in \mathbb{N} - \mathbb{N}$, let $M_{\eta} = \{\inf_{1 \le i \le \eta} Y_i < -m_0\}$. Then

$$\int_{*X} *Y_1 d*P + m_0 *P(M_\eta) \leq \int_{*X-M_\eta} Y_\eta d*P.$$

Since both integrals are finite and m_0 is infinite, $*P(M_{\eta}) \cong 0$. Since η is arbitrary in *N - N, $*P(\inf_{i \in *J} Y_i < -m_0) \cong 0$, a contradiction.

Assume $\{Y_i\}$ is an L^1 -bounded martingale, i.e., both $\{Y_i\}$ and $\{-Y_i\}$ are submartingales, and assume the submartingale $\{|Y_i|\}$ is uniformly integrable. Then it is a well-known and now easily obtained fact that for $Z = \lim Y_i$, we have $Y_i = E[Z | \mathcal{F}_i]$ a.e. for each *i*.

References

1. V. Alda, On conditional expectations, Chech. Math. J. 5 (1955), 503-505.

2. J. L. Doob, Stochastic Processes, Wiley, New York, 1953.

3. R. Hersh and P. Greenwood, *Stochastic differential and quasi-standard random variables*, in *Conference on Probabilistic Methods and Differential Equations*, E. A. Pinsky, ed., Lecture Notes in Math. No 451. Springer-Verlag, Berlin and New York, (1975), 35–61.

4. P. A. Loeb, An introduction to nonstandard analysis and hyperfinite probability theory, in Probabilistic Analysis and Related Topics, Vol. 2, A. T. Bharucha-Reid, ed., Academic Press, New York, 1979, 105–142.

5. A. Robinson, Non-standard Analysis, North-Holland Publ., Amsterdam, 1966.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS, URBANA, IL 61801