

Reprinted from *The Australian Journal of Statistics*, Vol. 14, No. 1,
pages 1–9, April, 1972

**MARTINGALES: A CASE FOR A PLACE IN THE
STATISTICIAN'S REPERTOIRE**

—
C. C. HEYDE
Australian National University

MARTINGALES : A CASE FOR A PLACE IN THE STATISTICIAN'S REPERTOIRE¹

C. C. HEYDE

Australian National University

1. Introduction

It is usually the case that martingale methods do not form a valued part of the statistician's repertoire of techniques. The reasons are partly historical and partly environmental, and while modern texts on probability theory deal in great detail with sums of independent random variables, the essentially broader subject of martingales is often relegated to a small section or the back pages. The object of this note is to give a brief exposition of what martingales are and what they can do. It is my personal feeling that the rôle of martingale theory in probability and statistics will increase rapidly in importance within the next few years.

2. The Basic Ideas

In order to give a feel for the subject, we start with the basic concepts in the case of a countable state space.

Let Ω be a sample space consisting of countably many points $\omega_1, \omega_2, \dots$ to which are assigned probabilities p_1, p_2, \dots respectively, with $p_j \geq 0$ and $\sum_j p_j = 1$. If X is a random variable its *expectation* is defined as

$$EX = \sum_j X(\omega_j)p_j$$

(provided $\sum_j |X(\omega_j)|p_j < \infty$) and the *conditional expectation* of X relative to A is defined correspondingly, when $P(A) > 0$, as

$$E(X | A) = \sum_{\{j: \omega_j \in A\}} X(\omega_j)p_j/P(A).$$

If $P(A) = 0$, the conditional expectation can be defined arbitrarily.

Now let $\{A_n, n \geq 1\}$ be a partition of Ω , that is, a countable class of disjoint sets with union Ω . This partition generates, and is in turn determined by, a σ -field, \mathcal{F} , namely the class of all unions of sets of the partition. If X is a random variable with an expectation, define $E(X | \mathcal{F})$, the *conditional expectation* of X relative to \mathcal{F} , as the random variable with constant value $E(X | A_n)$ on each set A_n . The definition specifies $E(X | \mathcal{F})$ almost surely (a.s.), that is, everywhere except on partition sets (if any) of probability zero. As an example, if Y_1, \dots, Y_k are random variables, they induce a partition each of whose sets is determined by a condition of the form

$$\{Y_1 = a_1, \dots, Y_k = a_k\}.$$

Let \mathcal{F}_k be the σ -field generated by this partition. In this case $E(X | \mathcal{F}_k)$ (which is commonly written as $E(X | Y_1, \dots, Y_k)$) is the random variable with value $E(X | Y_1 = a_1, \dots, Y_k = a_k)$ on the set $\{Y_1 = a_1, \dots, Y_k = a_k\}$.

¹ Manuscript received February 22, 1972.

Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$ be an increasing sequence of σ -fields generated by partitions of the sample space as just described. It is useful to interpret \mathcal{F}_n as representing the class of all relevant past events up to and including time n . The monotoneity relation, then, corresponds to the idea that the past up to time $n+1$ includes more events than the past up to time n . Let X_1, X_2, \dots be a sequence of random variables such that X_n is determined by the history up to and including time n (i.e. each event of the form $\{X_n=a\} \in \mathcal{F}_n$; X_n is then called \mathcal{F}_n measurable). Then, the sequence $\{X_n, n \geq 1\}$ is defined as a *martingale* (MG) relative to $\{\mathcal{F}_n, n \geq 1\}$ if each X_n has finite expectation and if for $m < n$,

$$(1) \quad E(X_n | \mathcal{F}_m) = X_m \text{ a.s.}$$

This is a relation between functions on the sample space and is to hold almost surely (a.s.), that is, everywhere except perhaps on a subset of the sample space of probability zero. Note that it is sufficient to replace (1) by

$$E(X_{n+1} | \mathcal{F}_n) = X_n \text{ a.s., each } n \geq 1,$$

for then

$$\begin{aligned} E(X_n | \mathcal{F}_m) &= E\{E(X_n | \mathcal{F}_{n-1}) | \mathcal{F}_m\} \\ &= E(X_{n-1} | \mathcal{F}_m) \\ &= \dots = E(X_m | \mathcal{F}_m) = X_m \text{ a.s.} \end{aligned}$$

using continued reduction.

Now the basic definition of a MG is inspired by notions of fairness in gambling. The idea of a fair game is that the knowledge of the past should not enable the gambler to improve on his fortunes. As a simple example of the above idea let $Z_i = \pm 1$ each with probability $\frac{1}{2}$ and suppose that Z_1, Z_2, \dots are independent (i.e. basically a simple coin-tossing situation). Suppose the gambler's strategy is that he bets $b_n(Z_1, \dots, Z_n)$ on the $(n+1)$ st trial and let

$$S_n = S_n(b_1, \dots, b_n; Z_1, \dots, Z_n)$$

denote his fortune after n trials. $\{S_n\}$ satisfies the recurrence relation

$$S_{n+1} = S_n + Z_{n+1} b_n(Z_1, \dots, Z_n).$$

Then, if \mathcal{F}_n denotes the σ -field generated by Z_1, \dots, Z_n , we observe that $\{S_n, n \geq 1\}$ is a MG with respect to $\{\mathcal{F}_n, n \geq 1\}$. We have for $n \geq 1$,

$$\begin{aligned} E(S_{n+1} | \mathcal{F}_n) &= E(S_n + Z_{n+1} b_n(Z_1, \dots, Z_n) | \mathcal{F}_n) \\ &= S_n + b_n(Z_1, \dots, Z_n) E(Z_{n+1}) \\ &= S_n \end{aligned}$$

as required. Of course if $b_n(Z_1, \dots, Z_n) \equiv 1$, then S_n is just a sum of independent and identically distributed random variables.

Now suppose that the gambler can decide to skip individual trials. Introduce a decision function ε_n , which is \mathcal{F}_{n-1} measurable, such that ε_n is 0 or 1. If ε_n is 0 the gambler skips the n th trial and if $\varepsilon_n = 1$ he bets. Then, denoting his fortune after n trials by T_n , we have

$$T_{n+1} = T_n + \varepsilon_{n+1} Z_{n+1} b_n(Z_1, \dots, Z_n)$$

so that

$$\begin{aligned} E(T_{n+1} | \mathcal{F}_n) &= T_n + \varepsilon_{n+1} b_n(Z_1, \dots, Z_n) E(Z_{n+1}) \\ &= T_n \end{aligned}$$

and the game remains "fair" under this system of optional sampling.

Martingale theory forms a natural setting for many such gambling problems. It can be used to establish intuitively striking propositions such as *the persistent gambler goes broke with probability one* (Breiman [2], pp. 101, 102).

3. A General Definition

In many problems the assumption of the countability of the sample space Ω is not satisfied. We go on to sketch a general framework for a definition of a MG.

Let $\{\Omega, \mathcal{F}, P\}$ be a measure space: Ω is a set, \mathcal{F} is a σ -field of subsets of Ω and P a probability measure defined on \mathcal{F} . Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$ be an increasing sequence of σ -fields of \mathcal{F} sets. Let $\{X_n, n \geq 1\}$ be a sequence of functions on Ω satisfying

- (i) X_n is measurable relative to \mathcal{F}_n ,
- (ii) $E | X_n | < \infty$,
- (iii) If $m < n$, then $E(X_n | \mathcal{F}_m) = X_m$ a.s.

Then, the sequence of random variables is said to be a MG relative to $\{\mathcal{F}_n, n \geq 1\}$.

In many applications, the σ -field \mathcal{F}_n is the past as determined by X_1, \dots, X_m ; that is \mathcal{F}_n is generated by X_1, \dots, X_n . In this case the basic martingale condition (iii) is commonly written

$$E(X_n | X_m, \dots, X_1) = X_m \text{ a.s.}$$

The more flexible setting of the definition is chosen since, for example, one has often to replace the conditioning variables X_m, \dots, X_1 by some functions of them.

4. The Martingale Convergence Theorem

A key reason for the importance of the concept of a MG is provided by the following powerful theorem.

The MG convergence theorem. Let $\{X_n\}$ be a MG such that $\limsup_{n \rightarrow \infty} E | X_n | < \infty$. Then, there exists a random variable X such that $\lim_{n \rightarrow \infty} X_n = X$ a.s. and $E | X | \leq \liminf_{n \rightarrow \infty} E | X_n |$.

This is an existence theorem; it tells us nothing about the limit random variable save that it has a finite first absolute moment. The theorem seems rather unexpected a priori and it is an extremely powerful tool which has led to a number of interesting results for which it seems essentially a unique method of approach. Of course one is still faced with finding the limit law, but that can usually be done by other methods.

A proof of the MG convergence theorem is particularly easy in the special case where $\limsup_{n \rightarrow \infty} EX_n^2 < \infty$. See for example Feller [6], p. 236. A general proof can be found in many of the standard probability texts.

As a simple example of the power of the theorem, consider its application to show that if $S_n = \sum_{i=1}^n X_i$ is a sum of independent random variables with S_n converging in distribution as $n \rightarrow \infty$, then S_n converges

a.s. This result is a straightforward consequence of the MG convergence theorem when it is noted that

$$Z_n = e^{itS_n} / E(e^{itS_n}) = e^{itS_n} / \prod_{i=1}^n E(e^{itX_i})$$

is a MG. The result is rather difficult to prove by other methods.

Another interesting application of the MG convergence theorem concerns the likelihood ratio. Suppose X_1, X_2, \dots is a stochastic process such that the joint density of (X_1, \dots, X_n) is either p_n or q_n ; the problem is to decide which. Define Y_n by

$$Y_n = \frac{q_n(X_1, \dots, X_n)}{p_n(X_1, \dots, X_n)}.$$

Y_n is likely to be small or large according as the true density is p_n or q_n and the asymptotic behaviour of $\{Y_n\}$ is clearly of importance.

For simplicity we assume that the densities p_n are strictly positive and continuous (although it certainly suffices to assume $q_n(x_1, \dots, x_n) = 0$ whenever $p_n(x_1, \dots, x_n) = 0$; see Doob [5], p. 348, or Chow, Robbins and Siegmund [4], p. 12). Then, if the p_n are the true densities, the conditional density of X_{n+1} given X_1, \dots, X_n is $p_{n+1}(X_1, \dots, X_n, y) / p_n(X_1, \dots, X_n)$ and hence

$$\begin{aligned} E(Y_{n+1} | X_n, \dots, X_1) &= \int_{-\infty}^{\infty} \frac{q_{n+1}(X_1, \dots, X_n, y)}{p_{n+1}(X_1, \dots, X_n, y)} \cdot \frac{p_{n+1}(X_1, \dots, X_n, y)}{p_n(X_1, \dots, X_n)} dy \\ &= \frac{q_n(X_1, \dots, X_n)}{p_n(X_1, \dots, X_n)} = Y_n \end{aligned}$$

so that $\{Y_n, n \geq 1\}$ is a MG with respect to the sequence of σ -fields generated by $X_1, \dots, X_n, n \geq 1$. Clearly $Y_n \geq 0$ and $EY_n = 1$, so that the MG convergence theorem ensures the existence of a random variable Y , with $EY \leq 1$, given by $\lim_{n \rightarrow \infty} Y_n = Y$ a.s.

Under suitable hypotheses, detailed information can be obtained on Y . For example, suppose the p_n 's and q_n 's correspond to independent X_j 's with a common distribution. Then, it is not difficult to show that $Y = 0$ a.s. ([5], p. 349; [4], p. 19). This is a result of precisely the type desired.

5. Stopping

Martingale theory is closely related to problems of *stopping* and in particular *optimal stopping*. We shall just sketch these concepts here; a comprehensive discussion of the material in this section is given by Chow, Robbins and Siegmund [4].*

Let $\{\mathcal{F}_n, n \geq 1\}$ be an increasing sequence of σ -fields. A stopping time is by definition a random variable t such that

$$P(t < \infty) = 1$$

and

$$\{t = n\} \in \mathcal{F}_n \text{ for each } n.$$

* Reviewed in this issue.—EDITOR.

If $\{X_n, n \geq 1\}$ is a MG with respect to $\{\mathcal{F}_n, n \geq 1\}$ we may for concreteness interpret the MG as the values assumed by the gambler's fortune as he plays a sequence of "fair" games. A stopping time is a strategy by which the gambler decides to stop playing and X_t is his terminal fortune. The requirement $\{t=n\} \in \mathcal{F}_n$ is just that the gambler's decision to stop at time n depends only on his past experience. One interesting and important problem is whether

$$EX_t = EX_1,$$

i.e., whether the property of "fairness" is preserved under the stopping time t .

This theory provides a simple method of deriving the Wald identity of sequential analysis. Let Y_1, Y_2, \dots be independent and identically distributed, $S_n = \sum_{i=1}^n Y_i$, and assume that for some real

$\lambda \neq 0$, $\varphi(\lambda) = Ee^{\lambda Y_1}$ exists and that $\varphi(\lambda) \geq 1$. Then, if N is a stopping time for the sums S_1, S_2, \dots such that $|S_n|$ is uniformly bounded for $n \leq N$ and $EN < \infty$, we have

$$E[\{\varphi(\lambda)\}^{-N} e^{\lambda S_N}] = 1.$$

This identity is derived by noting that

$$X_n = [\varphi(\lambda)]^{-n} e^{\lambda S_n}$$

is a MG and showing that

$$EX_N = EX_1 = 1$$

(e.g. Breiman [2], p. 100).

Now let $\{X_n, n \geq 1\}$ be any stochastic sequence with X_n measurable with respect to \mathcal{F}_n and let t be any stopping time. Suppose also that $E|X_n| < \infty$ for each n . Define $V = \sup EX_t$ where the supremum is taken over all stopping times for which EX_t exists. Optimal stopping theory is concerned with the problems of existence and nature of an optimal stopping variable (i.e. one for which $EX_t = V$).

As an example, we mention the problem of testing a simple hypothesis against a simple alternative with independent and identically distributed observations of constant cost. It is well known that a Bayes solution to the problem is a Wald sequential probability ratio test. Now, the problem of minimizing the risk in this context can be re-stated as a problem on optimal stopping ([4], pp. 46–49) and the stopping variable appropriate to the sequential probability ratio test turns out to be optimal ([4], pp. 71, 105). Some variants are also given in [4].

6. Limit Theorems

Zero mean random walks and sums of independent random variables centred at their expectations are MG's. The standard theory for sums of independent random variables, however, continues to proceed in its own right. Limit results such as the laws of large numbers, central limit theorem and law of the iterated logarithm have been major achievements and have had a substantial impact. Now, for the most part, the standard theory for sums of independent random variables can be put into a more general MG setting. From the point of view of applications there is considerable advantage in doing this. We shall sketch the basic ideas.

Let $\{S_n, n \geq 1\}$ be a zero mean MG with respect to $\{\mathcal{F}_n, n \geq 1\}$ and write $S_n = \sum_{j=1}^n X_j$. The first thing to note is that many results on sums of independent random variables in fact require only orthogonality of the increments ($E(X_i X_j) = 0, i \neq j$) and that this property holds for MG's whose increments have finite variance. As a simple example we see that the so-called Chebyshev inequality continues to hold for MG's. By Markov's inequality we have for any $\varepsilon > 0$,

$$P(|S_n| \geq \varepsilon) \leq \varepsilon^{-2} E S_n^2,$$

while

$$E S_n^2 = \sum_{j=1}^n E X_j^2 + 2 \sum_{i>j} E X_i X_j = \sum_{j=1}^n E X_j^2$$

since for $i > j$,

$$E(X_i X_j) = E(X_j E(X_i | \mathcal{F}_{i-1})) = E(X_j \{E(S_i | \mathcal{F}_{i-1}) - S_{i-1}\}) = 0.$$

This inequality for MG's gives immediately a *weak law of large numbers*. In fact, for MG's whose increments have finite variance,

$$P(|S_n| \geq n\varepsilon) \leq n^{-2} \varepsilon^{-2} \sum_{j=1}^n E X_j^2,$$

so that $n^{-1} S_n \rightarrow 0$ in probability if $n^{-2} \sum_{j=1}^n E X_j^2 \rightarrow 0$.

In the general case of the weak law of large numbers, where the finite variance assumption is dropped, there is a well-known set of necessary and sufficient conditions for convergence in probability

of $n^{-1} S_n$, $S_n = \sum_{j=1}^n X_j$, denoting a sum of independent random variables.

In fact (Loève [11], p. 278) $n^{-1} S_n \rightarrow 0$ in probability if and only if

- (i) $\sum_{j=1}^n P(|X_j| \geq n) \rightarrow 0$
- (ii) $n^{-1} \sum_{j=1}^n E X_{jn} \rightarrow 0$
- (iii) $n^{-2} \sum_{j=1}^n [E X_{jn}^2 - (E X_{jn})^2] \rightarrow 0$

where X_{jn} is defined by $X_{jn} = X_j$ or 0 according as $|X_j| < n$ or $|X_j| \geq n$. An inspection of the proof given by Loève shows that the "if" part (which is after all the most important part) depends only on truncation and the use of the so-called Chebyshev inequality. A minor variant obtained by, in (ii), replacing $E X_{jn}$ by $E(X_{jn} | X_{j-1}, \dots, X_1)$ and $\rightarrow 0$ by $\rightarrow 0$ in probability and in (iii) replacing $(E X_{jn})^2$ by $E(E(X_{jn} | X_{j-1}, \dots, X_1))^2$, holds in the MG context. This of course reduces to the Loève condition in the special case of independence. The "only if" part of Loève's proof involves the use of characteristic functions and only this rests on the assumption of independence.

In the case of the *strong law of large numbers*, the most basic tool is probably the Kolmogorov criterion that if the X_i are independent

with zero mean and $EX_i^2 < \infty$ each i , then $\sum b_n^{-2} EX_n^2 < \infty$, $b_n \uparrow \infty$ entails $\lim_{n \rightarrow \infty} b_n^{-1} S_n = 0$ a.s. (e.g. Loève [11], p. 238). That this result continues to hold in the case when $\{S_n\}$ is a MG is a simple consequence of the MG convergence theorem (e.g. Feller [6], p. 238).

In the case of the *central limit theorem* there is the following result of Brown [3]. Let $\{S_n, \mathcal{F}_n\}$ denote a zero mean MG whose increments have finite variance. Write

$$S_n = \sum_{i=1}^n X_i, \quad V_n^2 = \sum_{j=1}^n E(X_j^2 | \mathcal{F}_{j-1}) \quad \text{and} \quad s_n^2 = EV_n^2 = ES_n^2.$$

If (“p” denotes convergence in probability)

$$s_n^{-2} V_n^2 \xrightarrow{p} 1$$

and

$$s_n^{-2} \sum_{j=1}^n E(X_j^2 I(|X_j| \geq \varepsilon s_n)) \rightarrow 0$$

as $n \rightarrow \infty$ for any $\varepsilon > 0$ ($I(\cdot)$ denotes the indicator function) then

$$\lim_{n \rightarrow \infty} P(s_n^{-1} S_n \leq x) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^x e^{-\frac{1}{2}u^2} du.$$

This result reduces to the standard Lindeberg-Feller result (e.g. [11], p. 280) in the case of independent random variables as the condition

$s_n^{-2} V_n^2 \xrightarrow{p} 1$ is then trivially satisfied. The proof is a little more delicate than that of the Lindeberg-Feller result but uses similar ideas (whose use in the MG context dates back to Lévy; see for example [5], p. 383).

The following *law of the iterated logarithm* has been obtained by Stout [12]. Let $\{S_n\}$ denote a stationary and ergodic zero mean MG whose increments have finite variance σ^2 . Then

$$n^{-1} S_n = \sigma \zeta(n) (2 \log \log n/n)^{\frac{1}{2}}$$

where $\zeta(n)$ has its set of limit points confined to $[-1, 1]$ and $\limsup_{n \rightarrow \infty} \zeta(n) = +1$ a.s., $\liminf_{n \rightarrow \infty} \zeta(n) = -1$ a.s. The law of the iterated

logarithm clearly provides detailed information on the rate of convergence in the strong law of large numbers ($\lim_{n \rightarrow \infty} n^{-1} S_n = 0$ a.s.). The

result contains the classical law of the iterated logarithm for sums of independent and identically distributed random variables (due to Hartman and Wintner).

7. Some Applications of the Limit Theory

In this section we shall mention some applications of the results of Section 6. The examples chosen involve problems amenable to treatment using MG limit theorems but not using the corresponding results for sums of independent random variables. Basically, the MG limit theory of Section 6 seems relevant to any context where conditional expectations, given the past, have a simple form. This

is largely a consequence of results such as if $\{Z_n, n \geq 1\}$ is a sequence of random variables with finite means, then

$$\left\{ \sum_{i=1}^n [Z_i - E(Z_i | Z_{i-1}, \dots, Z_1)] \right\}$$

is a MG relative to the sequence of σ -fields generated by $Z_i, i \leq n$.

Our first application deals with a time series context. Consider a stationary autoregression $\{x(n)\}$ of the form

$$(2) \quad \sum_{k=0}^q \beta(k) \{x(n-k) - \mu\} = \varepsilon(n); \quad \beta(0) = 1,$$

$$E\varepsilon(n) = 0, \quad E\{\varepsilon(m)\varepsilon(n)\} = 0, \quad m \neq n,$$

where we assume that

$$\sum_{k=0}^q \beta(k) z^k \neq 0, \quad |z| \leq 1.$$

We also suppose that $\{\varepsilon(n)\}$ is a stationary stochastic sequence with $E\varepsilon^2(1) = \sigma^2$ satisfying

$$(3) \quad E(\varepsilon(n) | \mathcal{F}_{n-1}) = 0 \text{ a.s., all } n,$$

where \mathcal{F}_n is the σ -field generated by $\varepsilon(m), m \leq n$ and

$$E(\varepsilon^2(n) | \mathcal{F}_{n-1}) = \sigma^2 \text{ a.s., all } n.$$

In this context, \mathcal{F}_n is also the σ -field generated by $x(m), m \leq n$ and

$$\varepsilon(n) = x(n) - E(x(n) | \mathcal{F}_{n-1})$$

so that (3) (i.e. the assumption that any $\left\{ \sum_{n=-r}^s \varepsilon(n) \right\}$ is a MG) is

equivalent to assuming that $E(x(n) | \mathcal{F}_{n-1})$ is the best linear predictor and the best linear predictor is the best predictor (both in the least squares sense). We are also assuming that $E(\varepsilon^2(n) | \mathcal{F}_{n-1})$, the prediction variance given the past, is a.s.-constant. Both these assumptions, and particularly the former, seem natural and intuitively satisfying.

Now the classical theory of inference for the process (2) requires that the $\varepsilon(n)$ be independent and identically distributed with zero mean and variance σ^2 (e.g. Hannan [7], Chapter VI). However, it has been shown by Hannan and Heyde [8], with proofs resting primarily on a law of large numbers and a central limit theorem for MG's, that the classical theory goes through unchanged under the present conditions. In fact, the results of [8] are considerably stronger and deal with the general purely non-deterministic process

$$x(n) - \mu = \sum_{j=0}^{\infty} \alpha(j) \varepsilon(n-j); \quad \sum_{j=0}^{\infty} \alpha^2(j) < \infty; \quad \alpha(0) = 1;$$

$$E\varepsilon(n) = 0, \quad E\{\varepsilon(m)\varepsilon(n)\} = 0, \quad m \neq n,$$

but we have specialized here for clarity of exposition. We shall sketch the relevant results.

Let $x(1), x(2), \dots, x(N)$ be a sample of N consecutive observations on the process $\{x(n)\}$ and define the autocovariances

$$c(j) = c(-j) = N^{-1} \sum_{n=1}^{N-j} (x(n) - \bar{x})(x(n+j) - \bar{x}), \quad j \geq 0,$$

where $\bar{x} = N^{-1} \sum_{j=1}^N x(j)$. The classical results, namely that $\beta(j)$ and σ^2 can be estimated through

$$\sum_{j=0}^q \hat{\beta}(j)c(k-j) = \delta_{ok} \hat{\sigma}^2, \quad k=0, 1, \dots, q$$

where δ_{ok} is the Kronecker delta, while μ can be estimated by \bar{x} , remain true in the present context. So does the Anderson and Walker [1] central limit theorem for the autocorrelations $r(j) = c(j)/c(0)$. Furthermore, Heyde [9] has used an iterated logarithm result for MG's to give detailed (almost sure) information on the error involved in the estimation of $\beta(j)$, σ^2 and μ . MG methods form a natural tool in this context since the conditional expectations take a simple and relevant form. Note that

$$\left\{ \sum_{n=k+1}^N (x(n-k) - \mu)\varepsilon(n), N > k \geq 1 \right\}, \left\{ \sum_{n=1}^N [(x(n) - \mu)\varepsilon(n) - \sigma^2], N \geq 1 \right\},$$

and $\left\{ \sum_{n=1}^N \varepsilon(n), N \geq 1 \right\}$ are all zero mean MG's with respect to $\{\mathcal{F}_N\}$.

Another setting where MG results of the style of Section 6 provide a rather natural tool is in the problem of estimating growth and immigration parameters in a subcritical branching process with immigration. Heyde and Seneta [10] have proved consistency and asymptotic normality of their recommended estimators using a law of large numbers and a central limit theorem for MG's. Again the setting is such that conditional expectations have a very simple form. In particular, if $\{X_n\}$ denotes a subcritical Galton-Watson branching process with immigration with mean of offspring distribution $m (< 1)$ and mean of immigration distribution λ , then

$$E(X_n | X_{n-1}, \dots, X_1) = mX_{n-1} + \lambda.$$

References

- [1] Anderson, T. W., and Walker, A. M. (1964). "On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process." *Ann. Math. Statist.*, 35, 1296-1303.
- [2] Breiman, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.
- [3] Brown, B. M. (1971). "Martingale central limit theorems." *Ann. Math. Statist.*, 42, 59-66.
- [4] Chow, Y. S., Robbins, H., and Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping*. Houghton-Mifflin, Boston.
- [5] Doob, J. L. (1953). *Stochastic Processes*. Wiley, New York.
- [6] Feller, W. (1966). *An Introduction to Probability Theory and its Applications*. Vol. II. Wiley, New York.
- [7] Hannan, E. J. (1970). *Multiple Time Series*. Wiley, New York.
- [8] Hannan, E. J., and Heyde, C. C. (1971). "On limit theorems for quadratic functions of discrete time series." Mimeographed (to appear in *Ann. Math. Statist.*).
- [9] Heyde, C. C. (1971). "An iterated logarithm result for martingales and its application in estimation theory for autoregressive processes." Mimeographed (to appear in *J. Applied Prob.*).
- [10] Heyde, C. C., and Seneta, E. (1971). "Estimation theory for growth and immigration rates in a multiplicative process." Mimeographed (to appear in *J. Applied Prob.*).
- [11] Loève, M. (1963). *Probability Theory*. Van Nostrand, Princeton.
- [12] Stout, W. F. (1970). "The Hartman-Wintner law of the iterated logarithm for martingales." *Ann. Math. Statist.*, 41, 2158-2160.