CPSC 550: Machine Learning II	2008/9 Term 2
Lecture 19 — Mar. 31, 2009	
Lecturer: Nando de Freitas	*Scribe: Kevin Swersky*

This lecture introduces the concept of Martingales, which will eventually be used to prove the convergence of stochastic approximation with dependent noise. Some examples of Martingales are given, and a measure-theoretic definition of conditional expectation is introduced.

19.1 Martingales

Definition 19.1. *A sequence of random variables $\{S_n\}$ is a* ***martingale*** *with respect to the sequence of random variables $\{X_n\}$ if for all $n \geq 1$:*

(i) $E(|S_n|) < \infty$

(ii) $E(S_{n+1}|X_{1:n}) = S_n$

Example 19.2. *Let $X_1, X_2, ...$ be a sequence of independent random variables with $E(|X_k|) < \infty$, $E(X_k) = 0\ \forall k$. Then $S_n = \sum_{i=1}^{n} X_i$ is a martingale with respect to the sequence $\{X_n\}$.*

Proof: Property (i) is satisfied since n is a finite number, and therefore S_n is a finite sum of finite numbers which means it is also finite. To show that property (ii) is satisfied, we use induction:

$$E(S_1|X_1) = E(X_1) = 0$$

Now assume $E(S_n|X_{1:n}) = S_n$, then:

$$\begin{aligned} E(S_{n+1}|X_{1:n}) &= E(X_{n+1} + S_n|X_{1:n}) \\ &= E(X_{n+1}|X_{1:n}) + E(S_n|X_{1:n}) \\ &= E(X_{n+1}) + E(S_n|X_{1:n}) \text{ by independence} \\ &= 0 + E(S_n|X_{1:n}) \\ &= S_n \end{aligned}$$

□

Example 19.3 ([1]). *Let $X_0, X_1, ...$ be a discrete Markov chain taking values on a countable space $\mathfrak{X}$ with transition matrix P. Suppose $\pi : \mathfrak{X} \to \mathbb{R}$ is a bounded function that satisfies $\sum_{j \in \mathfrak{X}} P_{ij}\pi(j) = \pi(i)\ \forall i \in \mathfrak{X}$.*
In other words, n is a time index and $X_n = i$ if the Markov chain takes the value i at time n, and π is a vector of probabilities for each value in $\mathfrak{X}$ that X_n could take. $\pi = [\pi(1), \pi(2), ..., \pi(i), ..., \pi(|\mathfrak{X}|)]^T$.
$S_n = \pi(X_n)$ is a martingale with respect to the sequence $\{X_n\}$ since:

$$\begin{aligned}
&E(S_{n+1}|X_{1:n}) \\
&= E(\pi(X_{n+1})|X_{1:n}) \\
&= E(\pi(X_{n+1})|X_n) && \textit{by the Markov property} \\
&= \sum_{j \in \mathfrak{X}} P_{X_n j}\pi(j) = \pi(X_n) = S_n
\end{aligned}$$

19.2 Conditional Expectation

Consider the random variables X, Z on a measurable space $(\Omega, \mathfrak{F}, P)$.
Let

$$\begin{cases} X \in \{x_1, x_2, ..., x_n\} \\ Z \in \{z_1, z_2, ..., z_n\} \end{cases} \tag{19.1}$$

Then the random variable $Y = E(X|Z)$ is defined as follows:
If $Z(\omega) = Z_j$, then $Y(\omega) = E(X|Z = z_j) \doteq y_j = \sum_i x_i P(X = x_i|Z = z_j)$

Remark:

- The σ-algebra over Z, $\mathcal{H} = \sigma(Z)$ consists of 2^n elements.
- Y is $\mathcal{H}$-measurable

Hence,

$$\begin{aligned}
E(YI_{Z=z_j}) &= \int I_{Z=z_j} y_j dP(Y, Z) \\
&= P(Z = z_j) \sum_i x_i P(X = X_i|Z = z_j) \\
&= E(XI_{Z=z_j})
\end{aligned}$$

That is, if $H_j \in \mathcal{H} = \{Z = z_j\}$ then $E(YI_{H_j}) = E(XI_{H_j})$, and since for each $H \in \mathcal{H}$, I_H is a sum of I_{H_j}'s: $E(YI_H) = E(XI_H)$

We illustrate this with an example to make it more clear.

Example 19.4. *Let* $A = \{B_i : 1 \leq i \leq n\}$.

$$\begin{aligned}
E(X|A) &\doteq E(XI_A) \\
&= \int \int XI_A(Z)dP(X,Z) \\
&= \int \int X \left(\sum_i I_{B_i}(Z) \right) dP(X,Z) \\
&= \sum_i E(XI_{B_i})
\end{aligned}$$

Theorem 19.5 (Kolmogorov Fundamental Theorem (1933)). *Give an random variable X on a measurable space $(\Omega, \mathfrak{F}, P)$, and a a sub-$\sigma$-algebra of $\mathfrak{F}$, $\mathcal{H}$. There exists a random variable Y such that:*

- Y *is* $\mathcal{H}$*-measurable*
- $E(|Y|) < \infty$
- $\forall H \in \mathcal{H}$, $\int_H Y dP = \int_H X dP$

Bibliography

[1] Williams, D.: Probability with Martingales. Cambridge Mathematical Textbooks. (1991)