Math-Stat-491-Fall2014-Notes-V

Hariharan Narayanan

November 18, 2014

Martingales

1 Introduction

Martingales were originally introduced into probability theory as a model for 'fair betting games'. Essentially we bet on events of known probability according to these known values and the payoff is also according to these values. Common sense tells us that in that case, in the long run, we should neither win nor lose. The theory more or less captures this idea but there are paradoxes - for instance if you play until you are ahead you will gain but you might need unbounded resources to reach that stage. However, the value of martingale theory far exceeds the original reason for their coming into being. These notes are intended to atleast partially justify this statement.

2 Preliminaries

The theory of martingales makes repeated use of the notion of conditional expectation. We review the theory briefly.

The conditional expectation of a random variable X given that a random variable $Y = y$, is given by

$$E[X \mid \{Y = y\}] \equiv \int x \times p_{X|Y}(x \mid y)dx.$$

In the discrete case this becomes

$$E[X \mid \{Y = y\}] \equiv \Sigma_i x_i \times p_{X|Y}(x_i \mid y).$$

Therefore we have

$$E[X] = \Sigma_i x_i \times p_X(x_i) = \Sigma_j[\Sigma_i x_i \times p_{X|Y}(x_i \mid y_j)] \times p_Y(y_j).$$

Now the expression $E[X \mid \{Y = y\}]$ defines a function of y. Consider the function $f(\cdot)$ on the sample space defined by

$$f(\omega) \equiv E[X \mid \{Y = Y(\omega)\}].$$

To compute the value of $f(\cdot)$ on a point ω in the sample space, we first compute $Y(\omega)$, then compute the expectation $E[X \mid \{Y = Y(\omega)\}]$. This random variable $f(\cdot)$ is usually denoted $E[X \mid Y]$. Similarly, the random variable $(E[X \mid Y])^2$ is simply the square of the random variable $E[X \mid Y]$.

Note that, while $X \mid \{Y = y\}$ is a random variable whose sample space is the set of all ω which map to y under the random variable Y, the expression $X \mid Y$ does not have a predetermined meaning.

Next, what meaning should we assign $Var[X \mid Y]$? Formally, $Var[Z] \equiv E[Z^2] - (E[Z])^2$. So we define

$$Var[X \mid Y] \equiv E[X^2 \mid Y] - (E[X \mid Y])^2.$$

Thus $Var[X \mid Y]$ is the random variable given by the expression on the right hand side.

We will now prove the *law of total expectation:*
$E[E[X \mid Y]]$ of the random variable $E[X \mid Y]$ is equal to $E[X]$. In the discrete case, we could first partition the sample space into preimages under Y of the values $Y(\omega) = y_i$, find the probability associated with each set, multiply it by $E[X \mid \{Y = y_i\}]$ and then compute the sum of all such products. Thus we get

$$E[E[X \mid Y]] = \Sigma_i E[X \mid \{Y = y_i\}] \times p_Y(y_i).$$

Hence

$$E[E[X \mid Y]] = \Sigma_i \Sigma_j x_j p_{X|Y}(x_j \mid y_i) \times p_Y(y_i).$$

This, by using definition of conditional probability becomes

$$E[E[X \mid Y]] = \Sigma_i \Sigma_j x_j p_{XY}(x_j, y_i).$$

Interchanging summation we get

$$E[E[X \mid Y]] = \Sigma_j \Sigma_i x_j p_{XY}(x_j, y_i).$$

We can rewrite this as

$$E[E[X \mid Y]] = \Sigma_j x_j [\Sigma_i p_{XY}(x_j, y_i)],$$

i.e., as

$$E[E[X \mid Y]] = \Sigma_j x_j p_X(x_j)] = E[X].$$

In martingale theory we also come across expressions of the kind $E[E[X \mid Y, Z] \mid Z]$ and a useful result is a form of the law of total expectation:

$$E[X|Z] = E[E[X \mid Y, Z] \mid Z].$$

We will call this the *second* form of the law of total expectation.

To prove this evaluate both sides for some value of Z, say $Z = z_1$. LHS $= E[X|Z = z_1]$. Now $(X \mid Z = z_1)$ is a random variable G defined on the subset A, say, of the sample space where $Z(\omega) = z_1$. If originally a subset B of A had a measure $\mu(B)$ now it has the measure $\mu(B)/\mu(A)$, on this new sample space. The RHS $= E[E[(X \mid Z = z_1) \mid Y]] = E[E[G \mid Y]]$. By the law of total probability, $E[G] = E[E[G \mid Y]]$, so that for $Z = z_1$, both sides are equal, proving the required result.

Another way of looking at this result is as follows. Let us call the event corresponding to $Y = y_1, Z = z_1$, in the sample space, $A(y_1, z_1)$. The event $A(\cdot, z_1)$, corresponding to $Z = z_1$ is $\bigcup_{y_i} A(y_i, z_1)$, the union being disjoint union of pairwise disjoint subsets. Let us call the event corresponding to $X = x_i, Z = z_1$ in the sample space, $B(x_i, z_1)$. Computing $E[X \mid Z = z_1] = \Sigma_{x_i} x_i p(x_i \mid z_1) = \Sigma_{x_i} x_i p(x_i, z_1)/p(z_1)$, is in words, computing the expectation over each $B(x_i, z_1)$, and then summing it over all X_i. We could compute instead $\Sigma_{x_i}\Sigma_{y_j}x_ip(x_i \mid y_j, z_1)p(y_j) = \Sigma_{x_i}\Sigma_{y_j}x_ip(x_i, y_j, z_1)/p(z_1)$. Here we are breaking $B(x_i, z_1)$ into the family of disjoint subsets $B(x_i, z_1) \cap A(y_j, z_1)$, computing the expectation over each of these smaller sets according to probability $p(x_i \mid y_j, z_1)$, multiplying by $p(y_j)$ and then summing the values. This latter computation corresponds to the RHS of the equation,

$$E[X|Z = z_1] = E[E[X \mid Y, Z = z_1] \mid Z = z_1].$$

2.1 Cautionary Remark

Martingale theory is about 'expectations'. For reasons of easy readability we will not state conditions such as 'if X is integrable '. Such a condition essentially means $E[|X|] < \infty$. All the results, in these notes on martingales, may be taken to be valid, unless otherwise stated, only when the expectation operation yields a finite value on $|f(X_1, \cdots, X_n)|$, where X_i are the random variables and $f(\cdot)$, the function under consideration.

3 Definition and elementary properties of martingales

Let $\mathcal{S} \equiv X_0, \cdots, X_n, \cdots$ be a sequence of random variables. A second sequence of random variables $\mathcal{M} \equiv M_0, \cdots, M_n, \cdots$ is said to be a *martingale* with respect to $\mathcal{S}$, if $E[M_{n+1} \mid X_0, \cdots, X_n] = M_n$. We will replace the sequence $X_0, \cdots, X_n$, by $\mathcal{F}_n$ for short so that the definition reads

$$E[M_{n+1} \mid \mathcal{F}_n] = M_n.$$

1. (**Expectation remains invariant with time**)
   Applying 'expectation' operation on both sides of the defining equation of a martingale and using the law of total probability, we have

$$E[M_{n+1}] = E[E[M_{n+1} \mid \mathcal{F}_n]] = E[M_n].$$

2. (**In the martingale definition M_{n+k} can replace M_{n+1}**)
   Next, we note that $\mathcal{F}_{n+k} = \mathcal{F}_n, X_{n+1}, \cdots, X_{n+k}$.Using the *second* form of the law of total probability, we have $E[M_{n+k} \mid \mathcal{F}_n] = E[E[M_{n+k} \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n]$. Now $E[M_{m+r} \mid \mathcal{F}_m] = M_m, r = 1$. Suppose it is true that $E[M_{m+r} \mid \mathcal{F}_m] = M_m, r < k, \forall m$, then we must have $E[M_{n+k} \mid \mathcal{F}_{n+1}] = M_{n+1}$. We therefore have,

$$E[M_{n+k} \mid \mathcal{F}_n] = E[E[M_{n+k} \mid \mathcal{F}_{n+1}] \mid \mathcal{F}_n] = E[M_{n+1} \mid \mathcal{F}_n] = M_n.$$

   Further, by the law of total probability

$$E[M_{n+k}] = E[E[M_{n+k} \mid \mathcal{F}_n]] = E[M_n].$$

3. (**Orthogonality of martingale increments**)
   Let $j \leq k \leq l$. We then have

$$E[M_j(M_l - M_k)] = E[E[M_j(M_l - M_k)]|\mathcal{F}_j]] = E[M_jE[(M_l - M_k)|\mathcal{F}_j]] = E[M_j(M_j - M_j)] = 0.$$

Next let $i \leq j \leq k \leq l$. We then have

$$E[(M_j - M_i)(M_l - M_k)] = E[E[(M_j - M_i)(M_l - M_k)]|\mathcal{F}_j]] = E[(M_j - M_i)E[(M_l - M_k)|\mathcal{F}_j]]$$

$$= E[(M_j - M_i)(M_j - M_j)] = 0.$$

Note that this happens even though $(M_j - M_i), (M_l - M_k)$, are not independent random variables.

4. (**Variance consequence of orthogonality**)

$$Var[M_{m+k} - M_m] = E[(M_{m+k} - M_m)^2] - (E[M_{m+k} - M_m])^2.$$

$$E[(M_{m+k} - M_m)^2] = E[(M_{m+k})^2 + (M_m)^2] - 2E[(M_{m+k})(M_m)]$$

$$= E[(M_{m+k})^2 + (M_m)^2] - 2E[(M_m)(M_m)] = E[(M_{m+k})^2] + E[(M_m)^2] - 2E[(M_m)^2]$$

$$= E[(M_{m+k})^2] - E[(M_m)^2].$$

$$(E[M_{m+k} - M_m])^2 = (E[M_{m+k}] - E[M_m])^2 = (E[M_{m+k}])^2 + (E[M_m])^2 - 2(E[M_{m+k}])(E[M_m])$$

$$= (E[M_{m+k}])^2 + (E[M_m])^2 - 2(E[M_m])(E[M_m]) = (E[M_{m+k}])^2 - (E[M_m])^2.$$

So

$$Var[M_{m+k} - M_m] = Var[M_{m+k}] - Var[M_m].$$

Next since $Var[M_2 - M_1] = Var[M_2] - Var[M_1]$, it follows that $Var[M_2] = Var[M_1] + Var[M_2 - M_1]$. Similarly,

$$Var[M_n] = Var[M_1] + \Sigma_2^n Var[M_i - M_{i-1}].$$

4 Examples of Martingales

(a) (**Generalized random walk**)

Let $\mathcal{S} \equiv X_0, \cdots, X_n, \cdots$ be a sequence of independent random variables with common mean μ. Let $M_n \equiv \Sigma_0^n X_i - n\mu$. As before, let $\mathcal{F}_n$ denote $X_0, \cdots, X_n$. We then have,

$$E[M_{n+1} \mid \mathcal{F}_n] = E[M_n + X_{n+1} - \mu \mid \mathcal{F}_n] = M_n + E[X_{n+1} - \mu \mid \mathcal{F}_n] = M_n,$$

making M_n a martingale with respect to $\mathcal{S}$.

(b) (**Products of independent random variables**)

Let $\mathcal{S} \equiv X_0, \cdots, X_n, \cdots$ be a sequence of independent random variables with common mean μ, with $\mathcal{F}_n$ defined as before. Let $M_n \equiv (\mu)^{-n}\Pi_0^n X_i$. Then

$$E[M_{n+1} \mid \mathcal{F}_n] = E[(\mu)^{-(n+1)}\Pi_0^{n+1} X_i \mid \mathcal{F}_n]$$

$$= (E[(\mu)^{-1} X_{n+1} \mid \mathcal{F}_n])(\mu)^{-n}\Pi_0^n X_i = (\mu)^{-n}\Pi_0^n X_i = M_n.$$

(c) (**Branching processes**)

We know that in the case of these processes we have $\mu X_n = E[X_{n+1} \mid X_n] = E[X_{n+1} \mid \mathcal{F}_n]$. If we take $M_n \equiv \mu^{-n} X_n$, then we find

$$E[M_{n+1} \mid \mathcal{F}_n] = E[\mu^{-(n+1)} X_{n+1} \mid X_n] = \mu^{-n} X_n = M_n.$$

(d) (**Martingales for Markov chains**)

Suppose $\mathcal{S} \equiv X_0, \cdots, X_n, \cdots$ is a Markov chain. One way of associating a fair game with the Markov chain is to bet on the next state when the chain is at a state i. The transition matrix would be available to the bettor. The pay off would be $1/p(i,j), p(i,j) > 0$, if one bets 1 dollar on j. The expected gain would be 0, since the game is fair. One could define a martingale M_n for this situation as a function $f(i,n)$ of the present state i and the time n, i.e., $M_n \equiv f(i,n)$. In order that M_n becomes a martingale we need $E[M_{n+1} \mid \mathcal{F}_n] = M_n$. Now $E[M_{n+1} \mid \mathcal{F}_n] = \Sigma_j p(i,j) f(j, n+1)$, and $M_n = f(i,n)$. Therefore we must have $f(i,n) = \Sigma_j p(i,j) f(j,n+1)$, as a necessary condition for $M_n \equiv f(i,n)$ to be a martingale.

On the other hand if the function $f(i,n)$ satisfies $f(i,n) = \Sigma_j p(i,j) f(j,n+1)$, the above argument shows that $M_n \equiv f(i,n)$ is a martingale. Thus the condition

$$f(i,n) = \Sigma_j p(i,j) f(j,n+1), \qquad (*)$$

is necessary and sufficient for $M_n \equiv f(i,n)$ to be a martingale on the Markov chain. We give below a couple of instances of such functions.

i. Consider the 'gamber's ruin' Markov chain with $p(i,i+1) = p, p(i,i-1) = 1-p$. It can be verified that $f(i,n) \equiv ((1-p)/p)^i$ satisfies the condition (*) above so that $M_n \equiv f(i,n)$ is a martingale.

ii. When $p = 1/2$ in the gambler's ruin example another possible $f(i,n)$ is the function $i^2 - n$.

5 Optional stopping theorem

We state this important result without proof.

First we define stopping time for a stochastic process $\mathcal{S} \equiv X_0, \cdots, X_n, \cdots$. The positive integer valued, possibly infinite, random variable N, is said to be a *random time* for $\mathcal{S}$, if the event $\{N = n\}$ is determined by the random variables $X_0, \cdots, X_n$, i.e., if we know the values of $X_0, \cdots, X_n$, we can say whether $\{N = n\}$ is true or not. If $Pr\{N < \infty\} = 1$, then the random time N is said to be a *stopping time.*

Theorem 5.1. *(***Martingale Stopping Theorem***)*
Let $\mathcal{S} \equiv X_0, \cdots, X_n, \cdots$ and let M_n be a a martingale for $\mathcal{S}$. Let N be a random time for the process $\mathcal{S}$. If

- *N is bounded or*
- *$E[N] < \infty$ and there is a finite M such that $E[|M_{n+1} - M_n| \mid \mathcal{F}] < M$,*

then $E[M_N] = E[M_0]$.