University of Washington Math 523A Lecture 8

Lecturer: Eyal Lubetzky

Friday, April 24, 2009

1 Strong, weak, and very-weak martingales

Our goal will be to reach the following theorem, which we first state informally:

Theorem 1.1 (Kallenberg, Sztencel '91 ; Hayes '05). *Let $X = (X_n)_{n\geq 0}$ be a martingale in $\mathbb{R}^d$ with $X_0 = 0$. Then*

$$\mathbb{P}(\|X_n\| \geq a) \leq 2e^{1-\frac{(a-1)^2}{2n}} \quad \forall a > 0.$$

Notice the remarkable fact that the bound doesn't depend on the dimension d. We will see that it is sufficient to project the process to 2 dimensions.

What do we mean by a martingale in $\mathbb{R}^d$? We make the following definitions generalizing the notion of a martingale:

Definition 1.2 (Strong, weak, and very-weak martingales). *Suppose $X = (X_t)_{t\geq 0}$ is a process in $\mathbb{R}^d$ satisfying $\mathbb{E}\,\|X_t\| < \infty$ for all t (where $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^d$). We call X a* strong, weak, *or* very weak *martingale with respect to the induced filtration $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ if X satisfies one of the following conditions, respectively:*

1. *(strong) $\mathbb{E}[X_t \mid X_1, \ldots, X_{t-1}] = X_{t-1}$ for all t.*
2. *(weak) $\mathbb{E}[X_t \mid X_s] = X_s$ for all $s < t$.*
3. *(very weak) $\mathbb{E}[X_t \mid X_{t-1}] = X_{t-1}$ for all t.*

More generally, if $\mathcal{G} = (\mathcal{G}_t)_{t\geq 0}$ is an indexed collection of σ-fields (not necessarily increasing) and $X_t \in \mathcal{G}_t$, then X is adapted to the filtration $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ defined by $\mathcal{F}_t = \bigvee_{s\leq t} \mathcal{G}_s$. In this case we say that X is a strong, weak, or very weak martingale with respect to $\mathcal{G}$ if $\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] = X_{t-1}\ \forall t$, $\mathbb{E}[X_t \mid \mathcal{G}_s] = X_s\ \forall s < t$, or $\mathbb{E}[X_t \mid \mathcal{G}_{t-1}] = X_{t-1}\ \forall t$, respectively.

Note that strong $\Rightarrow$ weak $\Rightarrow$ very-weak, but the reverse implications don't necessarily hold. For example, if X is a very-weak martingale, our best guess for X_t given X_{t-1} is X_{t-1}, but we might be able to make a better guess if we're also given $X_1, \ldots, X_{t-2}$.

Recall: In the proof of the Hoeffding inequality, we only used the fact that X is a very-weak martingale (condition 3 above) since we always just used an increment of 1. This special case illustrates a general principle:

Theorem 1.3 (Hayes '05)**.** *Suppose (X_t) is a very-weak martingale in $\mathbb{R}^d$. Then there exists a strong martingale (Y_t) in $\mathbb{R}^d$ such that*

$$(Y_{t-1}, Y_t) \sim (X_{t-1}, X_t) \quad \forall t \geq 1.$$

This essentially means that we can apply any large deviation estimate for martingales to very-weak martingales as well, as long as the bound only involves the distribution of one variable (or two successive variables) at a time.

Proof sketch. We will prove the special case in which we have a finite number of steps and X_t takes finitely many possible values, i.e. $t \in \{1, \ldots, n\}$ for some finite n, and $X_t \in \Omega_t$, $|\Omega_t| < \infty \ \forall t$. We construct a coupling inductively as $X_1, X_2, \ldots$ are revealed to us. Let $Y_1 = X_1$, and for $t \geq 2$, assume $Y_1, \ldots, Y_{t-1}$ are already defined. Then we define the distribution of Y_t by

$$\mathbb{P}[(Y_0, \ldots, Y_t) = (y_0, \ldots, y_t)] = \begin{cases} 0, & \mathbb{P}(X_{t-1} = y_{t-1}) = 0 \\ \mathbb{P}(X_t = y_t \mid X_{t-1} = y_{t-1}) & \\ \quad \cdot \mathbb{P}[(Y_0, \ldots, Y_{t-1}) = (y_0, \ldots, y_{t-1})], & \text{otherwise.} \end{cases}$$

Equivalently,

$$\mathbb{P}(Y_t = y_t \mid Y_1 = y_1, \ldots, Y_{t-1} = y_{t-1}) = \mathbb{P}(X_t = y_t \mid X_{t-1} = y_{t-1}) \cdot \mathbf{1}_{\{\mathbb{P}(X_{t-1} = y_{t-1}) > 0\}}. \tag{1.1}$$

That is, to get Y_t, we "forget" how we got to Y_{t-1} and move according to X from Y_{t-1}. Then we have $(Y_{t-1}, Y_t) \sim (X_{t-1}, X_t)$ since we only used X_{t-1} and X_t in the definition, and by (1.1) we have

$$\mathbb{E}[Y_t \mid Y_1 = y_1, \ldots, Y_{t-1} = y_{t-1}] = \mathbb{E}[X_t \mid X_{t-1} = y_{t-1}] = y_{t-1}$$

since X is a very-weak martingale. This shows that

$$\mathbb{E}[Y_t \mid Y_1, \ldots, Y_{t-1}] = Y_{t-1},$$

so Y is a (strong) martingale.

The general case ($|\Omega_t| = \infty$, $n = \infty$) can be proved from the finite case above by using Kolmogorov's extension theorem to merge the finite-dimensional distributions together into a single measure. $\square$

We will now prove that, with regard to norms of martingales in $\mathbb{R}^d$, "Dimension 2 suffices."

Theorem 1.4 (Kallenberg, Sztencel 1991; Hayes '05)**.** *Let $X = (X_t)$ be a very-weak martingale in $\mathbb{R}^d$. Then there exists a strong martingale $Y = (Y_t)$ in $\mathbb{R}^2$ such that, for all t,*

$$\|Y_t\| \sim \|X_t\| \quad \text{and} \quad \|Y_t - Y_{t-1}\| \sim \|X_t - X_{t-1}\|.$$

Proof. Without loss of generality, assume $X_0 = 0$. We will couple X to a very weak martingale Y in $\mathbb{R}^2$ such that $\|Y_t\| = \|X_t\|$ and $\|Y_t - Y_{t-1}\| = \|X_t - X_{t-1}\|$. The result then follows from Hayes's Theorem 1.3 above.

Set $Y_0 = 0 \in \mathbb{R}^2$, and choose Y_1 to be one of the points $(\|X_1\|, 0)$ and $(-\|X_1\|, 0)$ in $\mathbb{R}^2$ with equal probability. For $t \geq 2$, suppose $Y_{t-1} \in \mathbb{R}^2$ has already been defined. There are two choices for $Y_t \in \mathbb{R}^2$ such that the triangles $\triangle(0, X_{t-1}, X_t) \subset \mathbb{R}^d$ and $\triangle(0, Y_{t-1}, Y_t) \subset \mathbb{R}^2$ are congruent – we use a fair coin flip ξ_t to decide between these two values, thus defining Y_t as a function of X_{t-1}, Y_{t-1}, X_t, and ξ_t. (We take the sequence $(\xi_t)_{t\geq 1}$ of coin flips to be IID and independent of X.) The congruence of the triangles means that $(\|Y_{t-1}\|, \|Y_t\|, \|Y_t - Y_{t-1}\|) = (\|X_{t-1}\|, \|X_t\|, \|X_t - X_{t-1}\|)$ for either choice of Y_t, so it remains to show that $Y = (Y_t)$ is a very weak martingale with respect to its induced filtration.

For all $t \geq 1$ we can decompose Y_t and X_t into components parallel to Y_{t-1} and X_{t-1}, respectively, and the orthogonal complements of these. That is, we have

$$\begin{aligned} Y_t &= \alpha_t Y_{t-1} + Z_t & \qquad\qquad Z_t &\perp Y_{t-1} \\ X_t &= \beta_t X_{t-1} + W_t & \qquad\qquad W_t &\perp X_{t-1} \end{aligned}$$

for some $\alpha_t, \beta_t \in \mathbb{R}$, and $Z_t \in \mathbb{R}^2$, $W_t \in \mathbb{R}^d$, where "$\perp$" denotes orthogonality with respect to the Euclidean inner product. More explicitly, we can take $\alpha_1 = \beta_1 = 1$ and $Z_1 = Y_1$, $W_1 = X_1$, and for each $t \geq 2$,

$$\begin{aligned} \alpha_t &:= \frac{\langle Y_t, Y_{t-1}\rangle}{\|Y_{t-1}\|^2} & \qquad\qquad Z_t &:= Y_t - \alpha_t Y_{t-1} \\ \beta_t &:= \frac{\langle X_t, X_{t-1}\rangle}{\|X_{t-1}\|^2} & \qquad\qquad W_t &:= X_t - \beta_t X_{t-1}, \end{aligned}$$

where $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product.

Since X is a very-weak martingale we have

$$\begin{aligned} X_{t-1} = \mathbb{E}[X_t \mid X_{t-1}] &= \mathbb{E}\left[\beta_t X_{t-1} + W_t \mid X_{t-1}\right] \\ &= \mathbb{E}[\beta_t \mid X_{t-1}] \cdot X_{t-1} + \mathbb{E}[W_t \mid X_{t-1}]. \end{aligned}$$

Since $W_t \perp X_{t-1}$ we have $\mathbb{E}[W_t \mid X_{t-1}] \perp X_{t-1}$, so this implies that $\mathbb{E}[W_t \mid X_{t-1}] = 0$ and $\mathbb{E}[\beta_t \mid X_{t-1}] = 1$.

Now, since $\triangle(0, X_{t-1}, X_t) \cong \triangle(0, Y_{t-1}, Y_t)$, we have $\alpha_t = \beta_t$ for all $t \geq 1$. Moreover, β_t depends only on X_t and X_{t-1}, so it is independent of $(\xi_t)_{t\geq 1}$ (which is independent of X). Therefore,

$$\mathbb{E}[\alpha_t \mid Y_{t-1}] = \mathbb{E}[\beta_t \mid X_{t-1}] = 1.$$

Finally, we have $\mathbb{E}[Z_t \mid Y_{t-1}] = 0$ since the two possible values of Z_t given Y_{t-1} are symmetric about the origin and occur with equal probability (according to ξ_t). Therefore,

$$\mathbb{E}[Y_t \mid Y_{t-1}] = \mathbb{E}[\alpha_t \mid Y_{t-1}] \cdot Y_{t-1} + \mathbb{E}[Z_t \mid Y_{t-1}] = Y_{t-1},$$

so Y is a very-weak martingale. We can now use Theorem 1.3 to construct a strong martingale from Y. $\square$

Remark: For the existence of the martingale Y in Theorem 1.4, "2 dimensions are required." For example, consider the martingale $X = (X_t)$ in $\mathbb{R}^2$ defined as follows: $X_t = \sum_{i=1}^t \xi_i$, where ξ_1 is chosen uniformly from the unit circle, and given X_{i-1} (for $i \geq 2$), ξ_i is chosen with equal probability from the two points which satisfy $\|\xi_i\| = 1$ and $\xi_i \perp X_{i-1}$ (these points are mirror images about the line through X_{t-1}). Then it follows by induction that $\|X_t\|^2 = t$ for all t; it is impossible to construct such a process in one dimension.

2 Maximal inequalities

Reminder (Optional Stopping Theorem): If (X_t) is a submartingale with respect to $\mathcal{F} = (\mathcal{F}_t)$ and $\tau_1 \leq \tau_2 \leq M < \infty$ are $\mathcal{F}$-stopping times, then $\mathbb{E}X_{\tau_1} \leq \mathbb{E}X_{\tau_2}$.

We will use this to prove the following maximal inequality:

Theorem 2.1. *Let $X = (X_t)$ be a submartingale with respect to $\mathcal{F} = (\mathcal{F}_t)$. Then for all integers n and all $x > 0$,*

$$\mathbb{P}\left(\max_{0\leq i\leq n} X_i \geq x\right) \leq \frac{\mathbb{E}\left[X_n \cdot \mathbf{1}_{\left\{\max_{0\leq i\leq n} X_i \geq x\right\}}\right]}{x} \leq \frac{\mathbb{E}X_n^+}{x}.$$

This result may look weak because the bound uses the first moment as in Markov's inequality (as opposed to the Chernoff or Hoeffding bounds which use infinite moments), but if n is very large, the maximal inequality can be much more powerful than the union bound.

Proof. let $\tau = \min\{i : X_i \geq x\}$. By optional stopping (for bounded stopping times), we have $\mathbb{E}X_{\tau\wedge n} \leq \mathbb{E}X_n$. In the case where $X_t > 0$ for all t, we have (using Markov's inequality):

$$\mathbb{P}\left(\max_{0\leq i\leq n} X_i \geq x\right) = \mathbb{P}(\tau \leq n) = \mathbb{P}(X_{\tau\wedge n} \geq x) \leq \frac{\mathbb{E}X_{\tau\wedge n}}{x} \leq \frac{X_n}{x}.$$

For the general case, we follow the proof of Markov's inequality:

$$\begin{aligned}\mathbb{E}X_n \geq \mathbb{E}X_{\tau\wedge n} &= \mathbb{E}\left[X_\tau \mathbf{1}_{\{\tau\leq n\}}\right] + \mathbb{E}\left[X_n \mathbf{1}_{\{\tau>n\}}\right] \\ &\geq x\mathbb{P}(\tau \leq n) + \mathbb{E}\left[X_n \mathbf{1}_{\{\tau>n\}}\right],\end{aligned}$$

so

$$x\mathbb{P}(\tau \leq n) \leq \mathbb{E}\left[X_n\left(1 - \mathbf{1}_{\{\tau>n\}}\right)\right] = \mathbb{E}\left[X_n \mathbf{1}_{\{\tau\leq n\}}\right].$$

□

For example, if (X_i) are IID, $\mathbb{E}X_1 = 0$, $\mathbb{E}X^2 < \infty$, and $S_n = \sum_{i=1}^n X_i$, then

$$\mathbb{P}\left(\max_{0\leq j\leq n} |S_j| \geq x\right) = \mathbb{P}\left(\max_{0\leq j\leq n} |S_j|^2 \geq x^2\right) \leq \frac{\operatorname{Var} S_n}{x^2}.$$

This is like Chebyshev's inequality, but for the maximum.