Survival Analysis
Week 3: Martingales

1 Filtrations and martingales

For a stochastic process $X(s)$ the filtration $\mathcal{F}_t^X = \sigma(X(s) : 0 \leq s \leq t)$ contains "the history" of the process. $\mathcal{F}_{t-}$ represents the data available just prior to time t. A process Z is *adapted* to $\mathcal{F}_t$ if $Z(t) \in \mathcal{F}_t$. Then $\mathcal{F}_t^Z \subset \mathcal{F}_t$.

A stochastic process $M(t)$, $t \in [0, \tau]$, is a *martingale* with respect to $\mathcal{F}_t$ if

(i) $M(t)$ is right-continuous with left-hand limits (*cadlag*)

(ii) $M(t)$ is adapted to $\mathcal{F}_t$

(iii) $E(|M(t)|) < \infty, \forall t$

(iv) $M(t)$ has the martingale property

$$E(M(t)|\mathcal{F}_s) = M(s), s \leq t \leq \tau$$

In all our applcations, $M(0) = 0$ such that $E(M(t)) = 0$.

Martingales have uncorrelated increments, i.e., for $0 \leq u \leq v \leq s \leq t$,

$$\begin{aligned}
&\operatorname{cov}\left(M(t) - M(s), M(v) - M(u)\right) \\
&\quad = E\left(\left(M(t) - M(s)\right)\left(M(v) - M(u)\right)\right) \\
&\quad = E\left(\underbrace{E\left(M(t) - M(s)|\,\mathcal{F}_v\right)}_{=M(v)-M(v)=0}\left(M(v) - M(u)\right)\right) = 0.
\end{aligned}$$

1.1 The Doob-Meyer decomposition

$\bar{M}(t)$ is a *submartingale* with respect to $\mathcal{F}_t$ if it satisfies $(i) - (iii)$ and

$$E(\bar{M}(t)|\mathcal{F}_s) \geq \bar{M}(s), s \leq t \leq \tau \tag{1}$$

The behaviour of a *predictable* process at t is determined by the information on $[0, t)$. If $X(t)$ is predictable, $X(t)$ is adapted to $\mathcal{F}_{t-}$. If $X(t) \in \mathcal{F}_t$ and is left-continuous, it is predictable. Any deterministic function is predictable. $X(t)$ is predictable if and only if X is adapted to the σ-algebra generated by all left-continuous adapted processes, see Fleming and Harrington (1991, Section 1.4).

A process $A(t)$ has *finite variation* if $\int_0^\tau |dA(s)| < \infty$. The difference between two nondecreasing processes has finite variation.

A is a *compensator* of the cadlag and adapted process X, if $X - A$ is a **mean zero martingale** and A is cadlag, predictable and has finite variation. If X has a compensator it is unique.

The Doob-Meyer decomposition: Any process $X(t) \in \mathcal{F}_t$ that is a difference of two submartingales has a compensator A. In particular, submartingales have compensators since the constant process 0 is a submartingale.

1.2 The predictable (co)variance

A *square integrable* martingale $(E(M(t)^2) < \infty,\ t \in [0, \tau])$ M satisfies the submartingale property, by Jensen's inequality,

$$E(M^2(t)|\mathcal{F}_s) \geq \left(E(M(t)|\mathcal{F}_s)\right)^2 = M^2(s).$$

As M^2 is a submartingale, it has a compensator $\langle M \rangle$, *the predictable variation process* of M. Since

$$M^2 - \langle M \rangle$$

is a mean zero martingale, the variance of $M(t)$ is given by

$$\mathrm{var}(M(t)) = E(M^2(t)) = E(\langle M \rangle(t)).$$

For M, $\tilde{M}$ square integrable martingales, $M\tilde{M}$ is a difference of submartingales. To see this note that $(M+\tilde{M})/2$ and $(M-\tilde{M})/2$ are martingales implying that $(M+\tilde{M})^2/4$ and $(M - \tilde{M})^2/4$ are submartingales, and

$$\begin{aligned}
&\frac{1}{4}(M + \tilde{M})^2 - \frac{1}{4}(M - \tilde{M})^2 \\
&\quad = \frac{M^2 + \tilde{M}^2 + 2M\tilde{M} - \left(M^2 + \tilde{M}^2 - 2M\tilde{M}\right)}{4} \\
&\quad = \frac{2M\tilde{M} + 2M\tilde{M}}{4} = M\tilde{M}
\end{aligned}$$

Because $M\tilde{M}$ is a difference of two submartingales, it has a compensator $\langle M, \tilde{M} \rangle$, the *predictable covariation process*, such that

$$M\tilde{M} - \langle M, \tilde{M} \rangle$$

is a mean zero martingale.

$\langle M, \tilde{M} \rangle$ is symmetric and bilinear like an ordinary covariance. Also,

$$\langle M_1 + M_2 \rangle = \langle M_1 \rangle + \langle M_2 \rangle + 2\langle M_1, M_2 \rangle$$

If $\langle M, \tilde{M} \rangle = 0$ we say that M and $\tilde{M}$ are orthogonal.

The quadratic covariation of M and $\tilde{M}$, is

$$[M, \tilde{M}](t) = \sum_{s \leq t} \Delta M(s) \Delta \tilde{M}(s).$$

The process

$$M\tilde{M} - [M, \tilde{M}]$$

is a martingale, but $[M, \tilde{M}]$ is not predictable. When $[M\tilde{M}]$ is square integrable,

$$[M\tilde{M}] - \langle M, \tilde{M} \rangle$$

is a martingale and $\langle M, \tilde{M} \rangle$ is the compensator of $[M\tilde{M}]$.

Note that

$$E([M](t)) = E(\langle M \rangle(t)) = \mathrm{var}(M(t))$$

and both $[M](t)$ and $\langle M \rangle(t)$ can be used for estimating the variance

1.3 The martingale transform

The integral of a bounded $\mathcal{F}_t$-predictable process $H(\cdot)$ with respect to a finite variation martingale,

$$\int_0^t H(s) dM(s),$$

is again a martingale, i.e., the martingale property is preserved under stochastic integration. The predictable and quadratic covariation processes are

$$\left\langle \int H dM, \int \tilde{H} d\tilde{M} \right\rangle = \int H\tilde{H} d \left\langle M, \tilde{M} \right\rangle$$
$$\left[\int H dM, \int \tilde{H} d\tilde{M} \right] = \int H\tilde{H} d \left[M, \tilde{M} \right].$$

Counting processes

Let T_i^*, $i = 1, \dots, n$, be i.i.d. absolute continous event times with hazard α. Let $\mathrm{A}(t) = \int_0^t \alpha(s) ds$. Let C_i be independent right-censoring times so that we only observe $T_i = T_i^* \wedge C_i$ and $\Delta_i = I\{T_i^* \le C_i\}$.

$$N_i(t) = I\{T_i \le t, \Delta_i = 1\}$$
$$N_i^C(t) = I\{T_i \le t, \Delta_i = 0\}$$

The filtration

$$\mathcal{F}_{t-} = \sigma(N_i(u), N_i^C(u) : 0 \le u < t, i = 1, \dots, n).$$

is the history on $[0, t)$. It contains $Y_i(t) = I\{T_i \ge t\}$, i.e., $Y_i(t) \in \mathcal{F}_{t-}$.

$N_i(t)$ is a submartingale, for $s < t$,

$$\begin{aligned} E(N_i(t)|\mathcal{F}_s) &= E((N_i(t) - N_i(s)) + N_i(s)|\mathcal{F}_s) \\ &= \underbrace{E(N_i(t) - N_i(s)|\mathcal{F}_s)}_{\ge 0} + N_i(s) \ge N_i(s). \end{aligned}$$

Thus, by the Doob-Meyer decomposition, $N_i(t)$, has a compensator Λ_i such that

$$M_i(t) = N_i(t) - \Lambda_i(t)$$

is a mean zero martingale and $E(N_i(t)) = E(\Lambda_i(t))$. $N_i(0) = 0$, so $E(\Lambda_i(0)) = E(N_i(0)) = 0$.

In the absolute continous case that we consider, there is a $\mathcal{F}_t$-predictable function $\lambda_i(s)$ called the *intensity* so that $\Lambda_i(t) = \int_0^t \lambda_i(s)ds$. Heuristically, from

$$E(dM_i(t)|\mathcal{F}_{t-}) = E(dN_i(t) - \lambda_i(t)dt|\mathcal{F}_{t-}) = 0$$

we have that $\lambda_i(t)dt = E(dN_i(t)|\mathcal{F}_{t-})$. With independent censoring, noting that $N_i(t)$ can only jump at t if $Y_i(t) = 1$,

$$d\Lambda_i(t) = E(dN_i(t)|\mathcal{F}_{t-}) = Y_i(t)dA(t) = Y_i(t)\alpha(t)dt.$$

For completeness, I supply the heuristic argument by verifying that $M_i(t) = N_i(t) - \int_0^t Y_i(u)\alpha(u)du$ fulfills the martingale property, for $0 \leq s \leq t \leq \tau$, $E(M_i(t)|\mathcal{F}_s) = M_i(s)$. From

$$\begin{aligned}
E(M_i(t)|\mathcal{F}_s) &= E\left(N_i(t) - \int_0^t Y_i(u)\alpha(u)du \middle| \mathcal{F}_s \right) \\
&= \underbrace{N_i(s) - \int_0^s Y_i(u)\alpha(u)du}_{=M_i(s)} \\
&\quad + E\left(\int_s^t dN_i(u) \middle| \mathcal{F}_s \right) - E\left(\int_s^t Y_i(u)\alpha(u)du \middle| \mathcal{F}_s \right). \qquad (2)
\end{aligned}$$

We need to show that (2) is zero. Consider the last term of (2). It can only be non-zero when neither N_i nor N_i^C has jumped before time s, so that

$$\begin{aligned}
E\left(\int_s^t Y_i(u)\alpha(u)du \middle| \mathcal{F}_s \right) &= Y_i(s-) \int_s^t E\left(Y_i(u) | T_i > s \right) \alpha(u)du \\
&= Y_i(s-) \frac{\int_s^t \mathrm{pr}(T_i \geq u)\alpha(u)du}{\mathrm{pr}(T_i > s)}. \qquad (3)
\end{aligned}$$

The first term in (2) is

$$\begin{aligned}
E\left(\int_s^t dN_i(u) \middle| \mathcal{F}_s \right) &= Y_i(s-)\mathrm{pr}\{s < T_i \leq t, \Delta_i = 1 | T_i > s\} \\
&= Y_i(s-) \frac{\mathrm{pr}\{s < T_i \leq t, \Delta_i = 1\}}{\mathrm{pr}(T_i > s)} \\
&= Y_i(s-) \frac{\mathrm{pr}(s < T_i^* \leq t, C_i \geq T_i^*)}{\mathrm{pr}(T_i > s)} \\
&= Y_i(s-) \frac{\int_s^t \left(-\frac{\partial}{\partial v} \mathrm{pr}(T_i^* \geq v, C_i \geq u)|_{v=u} \right) du}{\mathrm{pr}(T_i > s)} \\
&= Y_i(s-) \frac{\int_s^t \mathrm{pr}(T_i \geq u) \frac{-\frac{\partial}{\partial v} \mathrm{pr}(T_i^* \geq v, C_i \geq u)|_{v=u}}{\mathrm{pr}(T_i \geq u)} du}{\mathrm{pr}(T_i > s)} \qquad (4)
\end{aligned}$$

From the independent censoring assumption, i.e.,

$$\lim_{h\downarrow 0}\frac{1}{h}\mathrm{pr}(t \le T_i^* < t+h | T^* \ge t, C_i \ge t) = \lim_{h\downarrow 0}\frac{1}{h}\mathrm{pr}(t \le T_i^* < t+h | T_i^* \ge t) = \alpha(t),$$

we have that

$$\begin{aligned}\frac{-\frac{\partial}{\partial v}\,\mathrm{pr}(T_i^* \ge v, C_i \ge u)|_{v=u}}{\mathrm{pr}(T_i \ge u)} &= \frac{\lim_{h\downarrow 0}\frac{1}{h}\mathrm{pr}(u \le T_i^* < u+h, C_i \ge u)}{\mathrm{pr}(T_i^* \ge u, C_i \ge u)} \\ &= \lim_{h\downarrow 0}\frac{1}{h}\mathrm{pr}(u \le T_i^* < u+h, C_i \ge u | T_i^* \ge u, C_i \ge u) \\ &= \lim_{h\downarrow 0}\frac{1}{h}\mathrm{pr}(u \le T_i^* < u+h | T_i^* \ge u, C_i \ge u) \\ &= \lim_{h\downarrow 0}\frac{1}{h}\mathrm{pr}(u \le T_i^* < u+h | T_i^* \ge u) = \alpha(u). \qquad (5)\end{aligned}$$

Inserting (5) into (4), yields

$$Y_i(s-)\frac{\int_s^t \mathrm{pr}(T_i \ge u)\alpha(u)du}{\mathrm{pr}(T_i > s)}$$

which equals (3) and thus (2) is zero and the martingale property is established.

We now turn to the predictable and quadratic variation of M_i. As the compensator is continuous and doesn't jump, all jumps in the martingale M_i stem from jumps in N_i and the quadratic variation of M is

$$[M_i](t) = \sum_{s\le t}(\Delta M_i(s))^2 = \sum_{s\le t}(\Delta N_i(s))^2 = N_i(t).$$

As $\langle M_i\rangle$ is the compensator of $[M_i] = N_i$, and N_i has the *unique* compensator Λ_i,

$$\langle M_i\rangle(t) = \Lambda_i(t) = \int_0^t Y_i(s)\alpha(s)ds.$$

Assume that no two of the (continuous time) processes $N_1, \dots, N_n$ jump at the same time. Then,

$$[M_i, M_j](t) = \sum_{s<t}(\Delta M_i(s))(\Delta M_j(s)) = \sum_{s<t}(\Delta N_i(s))(\Delta N_j(s)) = 0, \text{ almost surely,}$$

$[M_i, M_j]$ has compensator $\langle M_i, M_j\rangle$, but since $[M_i, M_j] = 0$ we also have $\langle M_i, M_j\rangle = 0$ almost surely. The martingales M_i, $i = 1, \dots, n$, are orthogonal.

The sum of martingales

$$M_\bullet(t) = \sum_{i=1}^n M_i(t) = \sum_{i=1}^n \int_0^t Y_i(u)\alpha(s)ds = \int_0^t Y_\bullet(u)\alpha(s)ds,$$

where $Y_\bullet(t) = \sum_{i=1}^n Y_i(t)$, is a martingale with predictable variation

$$\begin{aligned}\langle M_\bullet\rangle(t) &= \sum_i \langle M_i\rangle(t) + 2\sum_{i<j}\overbrace{\langle M_i, M_j\rangle(t)}^{=0} \\ &= \sum_{i=1}^n \Lambda_i(t) = \int_0^t Y_\bullet(s)\alpha(s)ds\end{aligned}$$

The quadratic variation is

$$[M_\bullet](t) = \sum_{i=1}^n N_i(t) = N_\bullet(t).$$

For the martingales

$$\sum_{i=1}^n \int_0^t H_i(u) dM_i(u)$$

where $H_i(\cdot)$ are bounded and $\mathcal{F}_t$-predictable processes,

$$\left\langle \sum_{i=1}^n \int_0^\cdot H_i(u) dM_i(u) \right\rangle (t) = \sum_{i=1}^n \int_0^t H_i^2(u) d\langle M_i \rangle(u) = \sum_{i=1}^n \int_0^t H_i^2(u) Y_i(u) \alpha(u) du$$
$$\left[\sum_{i=1}^n \int_0^\cdot H_i(u) dM_i(u) \right] (t) = \sum_{i=1}^n \int_0^t H_i^2(u) d[M_i](u) = \sum_{i=1}^n \int_0^t H_i^2(u) dN_i(u).$$

The innovation theorem: Consider filtrations $\mathcal{G}_t$ and $\mathcal{H}_t$ such that $\mathcal{G}_t \subseteq \mathcal{H}_t$. If $N_i(t)$ is adapted to both filtratios and has intensity $\lambda_i^{\mathcal{H}}(t)$ with respect to $\mathcal{H}_t$, then the intensity with respect to $\mathcal{G}_t$ is

$$\lambda_i^{\mathcal{G}}(t) = E\left(\lambda_i^{\mathcal{H}}(t) \middle| \mathcal{G}_t \right).$$

A martingale CLT

Let U be a Gaussian martingale with $\langle U \rangle = [U] = V$, a right continuous deterministic function with positive semidefinite increments and $V(0) = 0$. This means that,

- $U(t) - U(s) \sim N(0, V(t) - V(s))$
- $U(t) - U(s) \perp\!\!\!\perp U(s)$

for all $u \leq s \leq t$.

For bounded and predictable H_i and square integrable martingales M_i, define

$$L^{(n)}(t) = \sum_{i=1}^n \int_0^t H_i(s) dM_i(s),$$
$$L_\epsilon^{(n)}(t) = \sum_{i=1}^n \int_0^t H_i(s) I\{|H_i(s)| > \epsilon\} dM_i(s)$$

that registers the jumps of size ϵ or more in $L^{(n)}$. The integrands in both processes are predictable, so both are martingales

CLT for martingales: If,

(a) $\langle L^{(n)} \rangle(t) \xrightarrow{P} V(t)$,

(b) *The Lindeberg conditon* $\langle L_\epsilon^{(n)} \rangle(t) \xrightarrow{P} 0$,

for all $t \in [0, \tau]$ and $\epsilon > 0$, then $L^{(n)} \xrightarrow{\mathcal{L}} U$ with variance process $V(t)$ and $\langle L^{(n)} \rangle(t) \xrightarrow{P} V(t)$ and $[L^{(n)}](t) \xrightarrow{P} V(t)$ uniformly.

For counting process martingales $M_i(t) = N_i(t) - \int_0^t Y_i(u)\alpha(s) ds$, the conditions are

(a) $\int_0^t H_i(s)^2 Y_i(s) \alpha(s) ds \xrightarrow{P} V(t)$,

(b) $\int_0^t H_i(s)^2 I\{|H_i(s)| > \epsilon\} Y_i(s) \alpha(s) ds \xrightarrow{P} 0$,

References

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* John Wiley & Sons.