BIO 244: Unit 12

Combining Martingales, Stochastic Integrals, and Applications to Logrank Test & Cox's Model

Because of Theorem 2.5.1 in Fleming and Harrington, see Unit 11: For counting process martingales with continuous compensators, the compensator fully determines the covariance function.

Example 12.1: Let $N(t) = 1(U \leq t, \delta = 1)$ and $A(t) = \int_0^t \lambda(u)\ Y(u)du$. Suppose that survival time T is continuous and that its hazard function $\lambda(\cdot)$ is bounded. Then $A(\cdot)$ is continuous and $E(M^2(t)) < \infty$ for all t. It follows that for $s \geq 0$, cov$(M(t), M(t+s))$=var$(M(t)) = \int_0^t \lambda(u)\ E(Y(u))du = \int_0^t \lambda(u)P(U \geq u)du$.

In this unit we will discuss functions of martingales. It is easy to show that a linear combination of martingales defined on the same filtration is also a martingale (see Exercises for Unit 11). **What about products of martingales?**

After considering this question, we will show that some stochastic integrals, say $Q(\cdot) = \int H(s)dM(s)$, with respect to a counting process martingale M are also martingales. These results will then be used to show that the numerator of the logrank test and Cox's score function can be represented as martingales processes. We conclude with a theorem that shows how to find the variance of $Q(t)$ as a simple function of the integrand $H(\cdot)$ and $M(\cdot)$.

We begin with a theorem that will enable us to evaluate the covariance between 2 martingale processes.

Theorem 12.1: Suppose that $M_1(\cdot)$ and $M_2(\cdot)$ are martingales defined on the same filtration, and that for every t, $E(M_j(t)) = 0$ and $E(M_j^2(t)) < \infty$, for j=1,2. Then there exists a right-continuous predictable process $< M_1, M_2 > (\cdot)$ such that
$M(\cdot) \stackrel{def}{=} M_1(\cdot)M_2(\cdot) - < M_1, M_2 > (\cdot)$ is a zero-mean martingale.

Proof: (adapted from F&H, Thm. 1.4.2). Note that $M_1(\cdot) + M_2(\cdot)$ and $M_1(\cdot) - M_2(\cdot)$ are zero-mean martingales, so that $(M_1(\cdot)+M_2(\cdot))^2$ and $(M_1(\cdot)-M_2(\cdot))^2$ are sub-martingales (via Jensen's inequality, see Unit 11 page 11, or directly). Thus, by the Doob-Meyer decomposition, there exist predictable, right-continuous $< M_1 + M_2, M_1 + M_2 > (\cdot)$ and $< M_1 - M_2, M_1 - M_2 > (\cdot)$ such that

$$M^{(M_1+M_2)}(\cdot) \overset{def}{=} (M_1(\cdot) + M_2(\cdot))^2 - < M_1 + M_2, M_1 + M_2 > (\cdot) \quad \textit{and} \tag{12.1}$$
$$M^{(M_1-M_2)}(\cdot) \overset{def}{=} (M_1(\cdot) - M_2(\cdot))^2 - < M_1 - M_2, M_1 - M_2 > (\cdot) \tag{12.2}$$

are martingales. We can choose these martingales to have zero mean. Now define

$$< M_1, M_2 > (\cdot) = \frac{< M_1+M_2, M_1+M_2 > (\cdot) - < M_1-M_2, M_1-M_2 > (\cdot)}{4}$$

and

$$M(\cdot) = M_1(\cdot)M_2(\cdot) - < M_1, M_2 > (\cdot).$$

Subtracting (12.2) from (12.1) and dividing by 4 yields

$$\frac{1}{4}\left(M^{(M_1+M_2)}(\cdot) - M^{(M_1-M_2)}(\cdot)\right) = \ldots = M_1(\cdot)M_2(\cdot) - < M_1, M_2 > (\cdot).$$

Since both of the terms on the left-hand side are zero-mean martingales defined on the same filtration, so is $M(\cdot)$.

Note: Note that this is not simply an application of the Doob-Meyer decomposition, since $M_1(\cdot)M_2(\cdot)$ is not in general a submartingale.

Note (compare with Exercise 6 Unit 10):

$$\text{cov}(M_1(t), M_2(t+s)) = E\left(< M_1, M_2 > (t)\right).$$

Thus, knowing $< M_1, M_2 >$ tells us about $\text{cov}(M_1(\cdot), M_2(\cdot))$.

Corollary 12.1: If $< M_1, M_2 > (\cdot) \stackrel{as}{=} 0$, then $M_1(\cdot)M_2(\cdot)$ is a martingale.

Definition: If $< M_1, M_2 > (\cdot) \stackrel{as}{=} 0$, then $M_1(\cdot)$ and $M_2(\cdot)$ are called orthogonal.

Note: This implies that orthogonal martingales are uncorrelated; i.e., because of the above, if $E(< M_1, M_2 > (t)) = 0$,

$$\text{cov}(M_1(t), M_2(t+s)) = E(< M_1, M_2 > (t)) = 0 \ .$$

In general, the process $M_1(\cdot)M_2(\cdot)$ is not necessarily a martingale. It will be when $M_1(\cdot)$ and $M_2(\cdot)$ are orthogonal. Thus, identifying martingales as orthogonal is useful.

The following theorem gives sufficient conditions for orthogonality.

Theorem 12.2: (F&H, Thm 2.5.2). Suppose that $N_1(\cdot),\ N_2(\cdot), ...,\ N_k(\cdot)$ are counting processes defined on the same filtration, with continuous compensators $A_1(\cdot),\ A_2(\cdot), ...,\ A_k(\cdot)$, respectively. Define $M_i(\cdot) = N_i(\cdot) - A_i(\cdot)$ for i=1,2,...,n. Then if no 2 of the counting processes can jump at the same time, $< M_i, M_j > (\cdot) \overset{a.s.}{=} 0$ for $i \neq j$. That is, $M_i(\cdot)$ and $M_j(\cdot)$ are orthogonal.

An obvious application of this result would be the observed survival experiences of k patients (if survival time is continuously distributed). Here our usual assumption of independence of these outcomes would ensure that no 2 of the patients' counting processes would jump at the same time.

We now consider an important result for processes formed as stochastic integrals with respect to counting process martingales.

Theorem 12.3 Let $N(\cdot)$ be a counting process with continuous compensator $A(\cdot)$, such that $M(\cdot) \stackrel{def}{=} N(\cdot) - A(\cdot)$ is a zero-mean martingale. Then if $H(\cdot)$ is any bounded, predictable process defined on the same filtration, the process $Q(\cdot)$ defined at time t by

$$Q(t) \stackrel{def}{=} \int_0^t H(s)dM(s)$$

is also a zero mean martingale.

Proof: See F & H, §1.5 for a proof. We do not include it here as it is somewhat complicated. However, one can get an intuitive feel for why the result is true by considering increasingly more difficult integrands $H(\cdot)$. For example, when $H(\cdot)$ is a constant in t (possibly stochastic), then it is clear that $Q(\cdot)$ is a martingale. Suppose next that $H(\cdot)$ is a step function with jumps at times $t_1 < t_2 < \cdots$. Then one can express the stochastic integral as a linear combination of increments, $M(t_{j+1}) - M(t_j)$, of the martingale $M(\cdot)$, in which case one would expect the result to also be a martingale. Or, with some hand-waving, for $u < t$:

$$\begin{aligned} E\left[\int_0^t H(s)dM(s)|\mathcal{F}_u\right] &\approx Q(u) + \sum_{u<s\leq t} E\left[H(s)dM(s)|\mathcal{F}_u\right] \\ &= Q(u) + \sum_{u<s\leq t} E\left[E\left[H(s)dM(s)|\mathcal{F}_{s-}\right]|\mathcal{F}_u\right] \\ &= Q(u) + \sum_{u<s\leq t} E\left[H(s)E\left[dM(s)|\mathcal{F}_{s-}\right]|\mathcal{F}_u\right] = 0, \end{aligned}$$

since "$H(s)$ is $\mathcal{F}_{s-}$-measurable" and where we have used the tower property of conditional expectations (see exercises).

To illustrate the value of this result, we give two examples.

Example 12.2 (Logrank Statistic): Assume the usual 2-sample setting, where

$$(U_i, \delta_i, Z_i), i = 1, 2, \ldots n$$

$$Z_i = \begin{cases} 0 & \text{if subject i is in group 0} \\ 1 & \text{if subject i is in group 1.} \end{cases}$$

Recall that the logrank test statistic is of the form

$$\frac{0 - E}{\sqrt{V}} = \frac{U}{\sqrt{V}}.$$

Consider U and define

$$Y_{ij}(u) = 1(U_i \geq u, Z_i = j)$$

and

$$N_{ij}(u) = 1(U_i \leq u, \delta_i = 1, Z_i = j) \quad \text{for } j = 0, 1 \; .$$

We earlier showed that (see page 3 of Unit 10)

$$U = \int_0^\infty \frac{Y_0(s)}{Y_0(s) + Y_1(s)} dN_1(s) - \int_0^\infty \frac{Y_1(s)}{Y_0(s) + Y_1(s)} dN_0(s),$$

where for j=0,1

$$Y_j(s) = \sum_{i=1}^n Y_{ij}(s) \qquad \text{and} \qquad N_j(s) = \sum_{i=1}^n N_{ij}(s).$$

Note that if

$$A_{ij}(t) \stackrel{def}{=} \int_0^t \lambda_j(u) Y_{ij}(u) du \quad j = 0, 1,$$

then

$$N_{ij}(t) - A_{ij}(t)$$

is a martingale and

$$dA_{ij}(u) = \lambda_j(u) Y_{ij}(u)\, du.$$

Thus we can write

$$\begin{aligned} U &= \sum_{i=1}^{n} \int \frac{Y_0(s)}{Y_0(s)+Y_1(s)} dN_{i1}(s) - \sum_{i=1}^{n} \int \frac{Y_1(s)}{Y_0(s)+Y_1(s)} dN_{i0}(s) \\ &= \sum_{i=1}^{n} \int \frac{Y_0(s)}{Y_0(s)+Y_1(s)} \left(dM_{i1}(s) + dA_{i1}(s)\right) \\ &\qquad - \sum_{i=1}^{n} \int \frac{Y_1(s)}{Y_0(s)+Y_1(s)} \left(dM_{i0}(s) + dA_{i0}(s)\right), \end{aligned}$$

where all integrals run from 0 to ∞. Under $H_0 : \lambda_0(\cdot) \equiv \lambda_1(\cdot)$, and it is easy to show that

$$\sum_{i=1}^{n} \int \frac{Y_0(s)}{Y_0(s)+Y_1(s)} dA_{i1}(s) - \sum_{i=1}^{n} \int \frac{Y_1(s)}{Y_0(s)+Y_1(s)} dA_{i0}(s) \stackrel{a.s.}{=} 0. \tag{12.3}$$

Thus under H_0,

$$U = \sum_{i=1}^{n} \int_0^{\infty} \left(\frac{Y_0(s)}{Y_0(s)+Y_1(s)} \right) dM_{i1}(s) - \sum_{i=1}^{n} \int_0^{\infty} \left(\frac{Y_1(s)}{Y_0(s)+Y_1(s)} \right) dM_{i0}(s).$$

Let $U(t)$ denote this expression when the integrals run from 0 to t (so that U(∞)=U), and consider the stochastic process $U(\cdot)$. Under H_0,

$$U(t) = \sum_{i=1}^{n} \int_0^{t} \left(\frac{Y_0(s)}{Y_0(s)+Y_1(s)} \right) dM_{i1}(s) - \sum_{i=1}^{n} \int_0^{t} \left(\frac{Y_1(s)}{Y_0(s)+Y_1(s)} \right) dM_{i0}(s).$$

Since $\frac{Y_0(\cdot)}{Y_0(\cdot)+Y_1(\cdot)}$ and $\frac{Y_1(\cdot)}{Y_0(\cdot)+Y_1(\cdot)}$ are bounded and predictable (since left continuous), each integral is a martingale (Theorem 12.3). Hence, also the sum of these integrals is a martingale, and so $U(t)$ is a martingale. Thus, *under the null*, the numerator of the logrank statistic can be viewed as the value of the martingale process U(t) at $t = \infty$. Looking ahead, we will show that the limit of this process (properly standardized) is Gaussian, and thus will be able to conclude that $U(\infty)$ has an asymptotic normal distribution, under the null.

Example 12.3 (PL score function from Cox's PH Model):

For simplicity (but without loss of generality) assume Z is a bounded scalar covariate. As before, the observations consist of

$$(U_i, \delta_i, Z_i) \qquad i = 1, 2, \ldots, n.$$

We assume the usual PH model in which $h(t \mid Z) = \lambda_0(t)e^{\beta Z}$, and that the data are continuous so that there are no ties. Then the partial likelihood score function (using the same notation as when we introduced Cox's model) is given by U=U(β), where (see Unit 7)

$$U = \sum_{\tau_j} \left(Z_{(j)} - \overline{Z}_j(\beta) \right),$$

with

$$\overline{Z}_j(\beta) = \frac{\sum_{\ell \in R_j} Z_\ell \; e^{\beta Z_\ell}}{\sum_{\ell \in R_j} e^{\beta Z_\ell}}.$$

This can also be expressed as (Exercise 7)

$$U = \ldots = \sum_{i=1}^{n} \int_0^\infty \left(Z_i - \frac{\sum_{\ell=1}^{n} Z_\ell Y_\ell(s) e^{\beta Z_\ell}}{\sum_{\ell=1}^{n} Y_\ell(s) e^{\beta Z_\ell}} \right) dN_i(s),$$

where $Y_i(s) = 1(U_i \geq s)$ and $N_i(s) = 1(U_i \leq s, \delta_i = 1)$.

Letting

$$A_i(s) \overset{def}{=} \int_0^s \left(\lambda_0(u)e^{\beta Z_i}\right) Y_i(u)du,$$

it follows that $M_i(s) \overset{def}{=} N_i(s) - A_i(s)$ is a zero-mean martingale and that

$$U(t) \overset{def}{=} \sum_{i=1}^{n} \int_0^t \left(Z_i - \frac{\displaystyle\sum_{\ell=1}^{n} Z_\ell Y_\ell(s) e^{\beta Z_\ell}}{\displaystyle\sum_{\ell=1}^{n} Y_\ell(s) e^{\beta Z_\ell}} \right) dN_i(s)$$

equals (Exercise 3)

$$U(t) = \sum_{i=1}^{n} \int_0^t \left(Z_i - \frac{\displaystyle\sum_{\ell=1}^{n} Z_\ell Y_\ell(s) e^{\beta Z_\ell}}{\displaystyle\sum_{\ell=1}^{n} Y_\ell(s) e^{\beta Z_\ell}} \right) dM_i(s). \tag{12.4}$$

Since the bracketed term is bounded (Exercise 8) and predictable (since left continuous), $U(t)$ is a martingale (Theorem 12.3). Thus, the score function from Cox's partial likelihood can be viewed as the value (when t=∞) of a martingale process.

The final result in this unit is a valuable theorem which tells us how to find the variance of a martingale formed as a stochastic integral with respect to a counting process martingale. A quick examination of the preceeding examples shows how it could be used to find the variance of the logrank statistic and Cox's partial likelihood score function.

Suppose that $N(\cdot)$ is a counting process and $A(\cdot)$ is its continuous compensator, so that $M(\cdot) \stackrel{def}{=} N(\cdot) - A(\cdot)$ is a zero-mean martingale. Then assuming $E\ M^2(t) < \infty$, we earlier showed that

$$\begin{aligned} \mathrm{Var}\,(M(t)) &= E\left(M^2(t)\right) = E\,(< M, M > (t)) \\ &= E\,(A(t))\,. \end{aligned}$$

This result was seen to be useful for getting the variance of counting process martingales.

Now consider the variance of the martingale

$$Q(t) = \int_0^t H(s)dM(s)$$

where $M(\cdot) = N(\cdot) - A(\cdot)$ and $H(\cdot)$ is a bounded and predictable process.

Theorem 12.4: (see F&H, Thm. 2.4.2). Assuming that $E(M^2(s)) < \infty$ for all s, and N is bounded, then for all t:

$$< Q, Q > (t) \stackrel{a.s.}{=} \int_0^t H^2(s)dA(s).$$

The proof is sketched in the Appendix. See F&H for a full proof. Under the conditions of this theorem, we can find the variance of Q. That is,

$$\begin{aligned} \text{Var}\,(Q(t)) &= E\left(Q^2(t)\right) \\ &= E\,(< Q, Q > (t)) = E\left(\int_0^t H^2(s)dA(s)\right). \end{aligned}$$

Thus, we can obtain the variance of

$$Q(t) = \int_0^t H(s)dM(s)$$

from knowledge of $H(\cdot)$ and $A(\cdot)$.

Note: Since the logrank numerator and Cox's score function can be expressed as sums of this type of stochastic integral, we can use this result to find their variances.

Exercises

1. Prove Corollary 12.1.

2. Verify equation (12.3).

3. Verify equation (12.4).

4. Prove that $Q(t)$ as defined above Theorem 12.4 has mean 0.

5. Assume that survival time is continuous and consider the 1-sample problem, where (U_i, δ_i), i=1,2,...,n are the observations. Define $N_i(t) = 1(U_i \leq t, \delta_i = 1)$ and $Y_i(t) = 1(U_i \geq t)$ for $t \geq 0$ and i=1,2,...,n.

   (a) Re-express the following Lebesgue-Stieltjes integral using the 'old' notation; that is, in terms of the numbers of persons who fail, are censored, or are at risk at various times.

   $$\sum_{i=1}^{n} \int_0^t \frac{Y_i(s)}{1+Y_i(s)} dN_i(s).$$

   (b) Suppose that $\lambda(t)$ denotes the hazard function for the underlying survival times. What is the variance of

   $$\sum_{i=1}^{n} \int_0^t \frac{Y_i(s)}{1+Y_i(s)} dM_i(s),$$

   where $A_i(\cdot)$ is the compensator process for $N_i(\cdot)$ and $M_i(\cdot) = N_i(\cdot) - A_i(\cdot)$? Express your answer in terms of $\lambda(\cdot)$ and the distribution function of the U_i.

6. Show that $EM^2(t) < \infty$ in Example 12.1.

7. Verify equation $U = ... = ...$ on page 9.

8. Show that the bracketed term in (12.4) is bounded.

9. Using the definition of conditional expectation, prove that if $\mathcal{F}_t \subset \mathcal{F}_s$, then $E[X|\mathcal{F}_t] = E[E[X|\mathcal{F}_s]|\mathcal{F}_t]$.

Appendix: Heuristic Proof that $< Q, Q > (t) = \int_0^t H^2(s)dA(s)$

We need to show that the unique compensator for $Q^2(t)$ is $\int_0^t H^2(s)dA(s)$; i.e., that $Q^2(t) - \int_0^t H^2(s)dA(s)$ is a martingale. The key thing to prove is thus that

$$E[Q^2(t+s) - \int_0^{t+s} H^2(u)dA(u) \mid \mathcal{F}_t] = Q^2(t) - \int_0^t H^2(u)dA(u) \ ,$$

or

$$E[Q^2(t+s) - \int_t^{t+s} H^2(u)dA(u) \mid \mathcal{F}_t] = Q^2(t) \ ,$$

or

$$E[Q^2(t+s) \mid \mathcal{F}_t] - E\left[\int_t^{t+s} H^2(u)dA(u) \Big| \mathcal{F}_t\right] = Q^2(t) \ ,$$

or that

$$E[Q^2(t+s) - Q^2(t) \mid \mathcal{F}_t] = E\left[\int_t^{t+s} H^2(u)dA(u) \Big| \mathcal{F}_t\right] . \qquad (A.1)$$

We show this heuristically by first taking $s = dt$ and showing that

$$E[Q^2(t+dt) - Q^2(t) \mid \mathcal{F}_t] = E[\int_t^{t+dt} H^2(u)dA(u) | \mathcal{F}_t],$$

or that, since A and H are predictable,

$$E[dQ^2(t) \mid \mathcal{F}_t] = E\left[d \int_0^t H^2(u)dA(u) \mid \mathcal{F}_t\right] = H^2(t)dA(t). \qquad (A.2)$$

By writing

$$\begin{aligned} dQ^2(t) &= Q(t+dt)^2 - Q^2(t) \\ &= (Q(t) + dQ(t))^2 - Q^2(t) \\ &= 2Q(t)dQ(t) + (dQ(t))^2, \end{aligned}$$

it follows that, since $Q(t)$ is $\mathcal{F}_t$-measurable and Q is a martingale,

$$\begin{aligned} E\left[dQ^2(t) \mid \mathcal{F}_t\right] &= 2E[dQ(t)\ Q(t)|\mathcal{F}_t] + E\left[(dQ(t))^2 \mid \mathcal{F}_t\right] \\ &= 2Q(t)E[dQ(t)|\mathcal{F}_t] + E\left[(dQ(t))^2 \mid \mathcal{F}_t\right] \\ &= 0 + E\left[(dQ(t))^2 \mid \mathcal{F}_t\right] . \end{aligned}$$

Thus, we need to show that

$$E\left[(dQ(t))^2 \mid \mathcal{F}_t\right] = H^2(t)dA(t).$$

But

$$(dQ(t))^2 = (H(t)dM(t))^2 = H^2(t)\ (dM(t))^2\ .$$

Thus,

$$\begin{aligned} E\left[(dQ(t))^2 \mid \mathcal{F}_t\right] &= E\left[H^2(t)(dM(t))^2 \mid \mathcal{F}_t\right] \\ &= H^2(t)\ E\left[(dM(t))^2 \mid \mathcal{F}_t\right] \\ &= H^2(t)\ E\left[dA(t) \mid \mathcal{F}_t\right] \\ &= H^2(t)dA(t). \end{aligned}$$

So much for taking $s = dt$. The heuristic proof is concluded by noticing that, with $I(t) = \int_0^t H^2(s)dA(s)$ and if $u > 0$,

$$\begin{aligned} &E\left[Q^2(t+u) - I(t+u)|\mathcal{F}_t\right] \\ &= Q^2(t) - I(t) + \sum_{t<s\leq t+u} E\left[dQ^2(s) - dI(s)|\mathcal{F}_t\right] \\ &= Q^2(t) - I(t) + \sum_{t<s\leq t+u} E\left[E\left[dQ^2(s) - dI(s)|\mathcal{F}_s\right]|\mathcal{F}_t\right] \\ &= Q^2(t) - I(t) + 0, \end{aligned}$$

where for the second equality we used the tower property of conditional expectations (see Exercises).