STK4080/9080 SURVIVAL AND EVENT HISTORY ANALYSIS

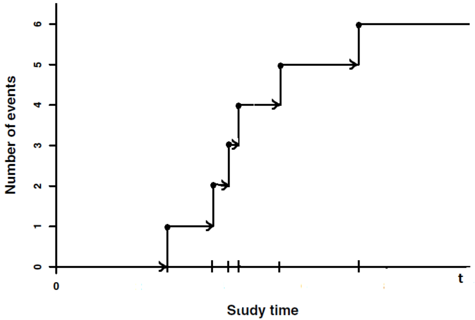
Slides 6: Counting processes

Bo Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim

https://www.ntnu.edu/employees/bo.lindqvist
bo.lindqvist@ntnu.no
boli@math.uio.no

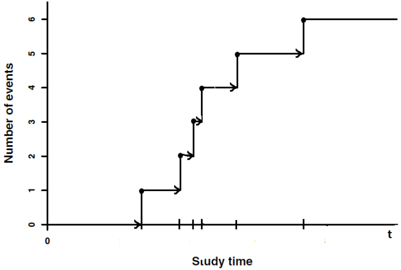
University of Oslo, Spring 2021

Counting Processes $N(t)$: Notation and basic facts (See section 1.4 in ABG)



$N(t) = \#$ events in $(0, t]$. *Some basic facts:*

- $N(t)$ makes jumps of size 1 at events
- $N(t)$ is right-continuous



Recall that we earlier defined $dM(t)$ as increase in $M(t)$ in the interval $[t, t + dt)$, which is $M((t + dt)-) - M(t-)$. For a counting process we can then define

$$dN(t) = \#\text{ events in } [t, t + dt) = \begin{cases} 1 & \text{if event at time } t \\ 0 & \text{if no event at time } t \end{cases}$$

We shall also write

$$N(t) = \int_0^t dN(s) \quad \text{and} \quad \int_0^t H(s)dN(s) = \sum_{i=1}^{k} H(t_i)$$

where $t_1, t_2, \ldots, t_k$ are the jump times of $N(t)$.

Poisson process $N(t)$

From course in stochastic processes, the Poisson process is such that

$$\begin{aligned} P(N(t+h) - N(t) = 1) &= \lambda h + o(h) \\ P(N(t+h) - N(t) \geq 2) &= o(h) \end{aligned}$$

as $h \to 0$, and $N(t)$ has **independent increments**. In slides 5 we wrote the last assumption to mean $N(t) - N(s)$ is independent of $\mathcal{F}_s$ for $s < t$.

With the new notation we can write

$$\begin{aligned} P(dN(t) = 1) &= P(\text{one event in } [t, t+dt)) = \lambda dt \\ P(dN(t) \geq 2) &= P(\text{two or more events in } [t, t+dt)) = 0 \end{aligned}$$

(the latter result is a general property we shall assume for counting processes)

Intensity process for a general point process

Recall for Poisson process,

$$P(dN(t) = 1) \equiv P(\text{event in } [t, t + dt)) = \lambda dt$$

For a point process, what happens in $[t, t + dt)$ may depend on the past before t, $\mathcal{F}_{t-}$. We therefore define the *intensity process* $\lambda(t)$ in general such that

$$P(dN(t) = 1|\mathcal{F}_{t-}) \equiv P(\text{event in } [t, t + dt) \mid \mathcal{F}_{t-}) = \lambda(t)dt$$

NOTE: This means that the intensity may be **random**; being a function of $N(s)$ for $s < t$. **AND**, since $\lambda(t)$ is known right before t, it is a **predictable** function.

BUT: For the Poisson process, what happens in $[t, \infty)$ is independent of $\mathcal{F}_{t-}$, so for the Poisson process, $\lambda(t) = \lambda$, i.e. a deterministic constant (which is trivially predictable).

*Non*homogenous Poisson process (NHPP) and extension

An NHPP assumes that λ may depend on t,

$$P(dN(t) = 1) = \lambda(t)dt$$

$\lambda(t)$ is then also the *intensity function* as defined on the previous slide.

The NHPP is muched used in reliability analyses of *repairable systems*, where the events are failure times of a technical system. The system is then assumed to be repaired in negligible time and restarted after each failure.

$\lambda(t)$ will reasonably be a decreasing function of time. But it may also happen that frequent repairs may decrease the intensity even more. A more general model for the intensity function has been suggested for failures of water pipes:

$$P(dN(t) = 1|\mathcal{F}_{t-}) \equiv \lambda(t)dt = (N(t-) + \beta)\alpha(t)dt$$

for a positive parameter β and a deterministic function $\alpha(t)$. The intensity is thus a random variable, and is seen to be a predictable process.

A single lifetime T modeled by a counting process

Let T be a lifetime with hazard function $\alpha(t)$. Then we can define the counting process

$$N(t) = I\{T \leq t\} \text{ for all } t \geq 0$$

in which case

$$P(dN(t) = 1 | \mathcal{F}_{t-}) \equiv P(t \leq T \leq t + dt | \mathcal{F}_{t-}) = \begin{cases} \alpha(t)dt & \text{for } T \geq t \\ 0 & \text{for } T < t \end{cases}$$

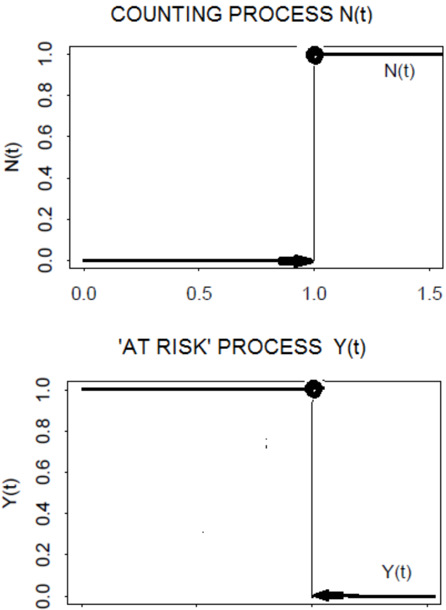
Thus the intensity function can be written

$$\lambda(t) = \alpha(t) I\{T \geq t\} \equiv \alpha(t) Y(t)$$

Here $Y(t) \equiv I\{T \geq t\}$ is called the **"at risk"** function. The $Y(t)$ is seen to be in $\mathcal{F}_{t-}$, since it is known right before T, and $Y(t)$ is furthermore **left-continuous**. See figure next slide.

NOTE THE CONCEPTUAL DIFFERENCE BETWEEN THE **HAZARD FUNCTION** $\alpha(t)$ AND THE **INTENSITY FUNCTION** $\lambda(t)$. EXPLAIN!

A single lifetime T modeled by a counting process



A single, possibly right censored, lifetime

Let T be a lifetime with hazard function $\alpha(t)$ and C a potential censoring time. Then the observations are $(\tilde{T}, D)$, where

$$\tilde{T} = \min(T, C), \quad D = I\{T = \tilde{T}\}$$

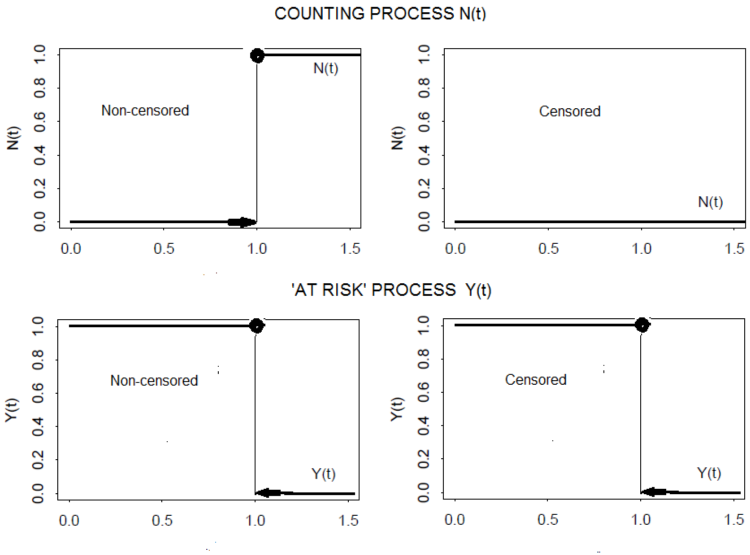
Define

- $N(t) = I\{\tilde{T} \leq t, D = 1\}$ (counting process)
- $Y(t) = I\{\tilde{T} \geq t\}$ ('at risk' process)

Again, the **intensity** is $\lambda(t) = \alpha(t)Y(t)$, which is a function of the history $\mathcal{F}_{t-}$.

See graphs on next slide

$$N(t) = I\{\tilde{T} \le t, D = 1\},\ Y(t) = I\{\tilde{T} \ge t\}$$



Right censored data

The standard setup for data from n individuals:

$$
\begin{aligned}
T_i &= \text{survival time for ind. } i \\
C_i &= \text{censoring time for ind. } i \\
\tilde{T}_i &= \min(T_i, C_i) = \text{observed time for ind. } i \\
\tilde{D}_i &= I(T_i = \tilde{T}_i) = \text{indicator for observed time for ind. } i
\end{aligned}
$$

Goal for statistical inference:

Estimate

$$
\begin{aligned}
\alpha_i(t) &= \text{hazard rate of } T_i \text{ (possibly the } \alpha_i(t) \text{ are equal)} \\
A_i(t) &= \int_0^t \alpha_i(s)ds = \text{cumulative hazard rate of } T_i \\
S_i(t) &= e^{-A_i(t)} = \text{survival function of } T_i
\end{aligned}
$$

The aggregated counting process

Let for $i = 1, \dots, n$,

$$\begin{aligned} N_i(t) &= I(\tilde{T}_i \le t, D_i = 1) = \textit{counting process} \\ Y_i(t) &= I(\tilde{T}_i \ge t) = \textit{'at risk' indicator} \\ \lambda_i(t) &= \alpha_i(t) Y_i(t) = \textit{intensity function} \end{aligned}$$

Then define the aggregated process:

$$\begin{aligned} N(t) &= \sum_{i=1}^{n} N_i(t) \\ \lambda(t) &= \sum_{i=1}^{n} \lambda_i(t) = \sum_{i=1}^{n} \alpha_i(t) Y_i(t) \end{aligned}$$

Now $N(t)$ satisfies the requirements for a counting process with intensity $\lambda(t)$. If the $\alpha_i(t) \equiv \alpha(t)$, then

$$\lambda(t) = \alpha(t) \sum_{i=1}^{n} Y_i(t) \equiv \alpha(t) Y(t)$$

which is known as **Aalen's multiplicative model**.

Independent censoring

Let $\mathcal{F}_t$ be the history of all individuals, their censorings and failures, i.e., $\mathcal{F}_t$ contains $\{(N_i(s), Y_i(s)), s \leq t, i = 1, \ldots, n\}$

Independent censoring means by definition (p. 30-31 in book):

$$\begin{aligned} P(t \leq \tilde{T}_i < t + dt, D_i = 1 | \tilde{T}_i \geq t, \mathcal{F}_{t-}) &= P(t \leq T_i < t + dt | T_i \geq t) \\ &= \alpha_i(t) dt \end{aligned}$$

ABG makes the point that

$$\textit{independent censoring} \Longleftrightarrow \lambda_i(t) = \alpha_i(t) Y_i(t) \quad (*)$$

where $\lambda_i(t)$ is the intensity of the observed process $N_i(t)$, recall

$$\lambda_i(t) dt =_{def} P(dN_i(t) = 1 | \mathcal{F}_{t-})$$

(see next page for an argument for $(*)$)

Proof of $(*)$

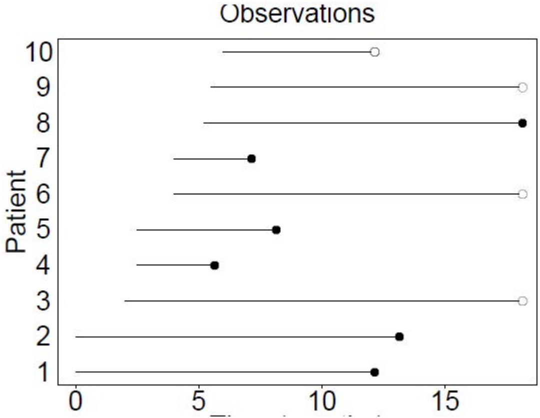
Claim: independent censoring $\iff \lambda_i(t) = \alpha_i(t) Y_i(t) \quad (*)$

$$
\begin{aligned}
\lambda_i(t)dt &= P(dN_i(t) = 1 | \mathcal{F}_{t-}) \\
&= P(t \le \tilde{T}_i < t + dt, D_i = 1 | \mathcal{F}_{t-}) \\
&= \textit{ using total probability rule:} \\
& \quad P(t \le \tilde{T}_i < t + dt, D_i = 1 | \tilde{T}_i \ge t, \mathcal{F}_{t-}) P(\tilde{T}_i \ge t | \mathcal{F}_{t-}) \\
&+ P(t \le \tilde{T}_i < t + dt, D_i = 1 | \tilde{T}_i < t, \mathcal{F}_{t-}) P(\tilde{T}_i < t | \mathcal{F}_{t-}) \\
&= P(t \le \tilde{T}_i < t + dt, D_i = 1 | \tilde{T}_i \ge t, \mathcal{F}_{t-}) P(\tilde{T}_i \ge t | \mathcal{F}_{t-}) + 0 \\
&= \textit{by def. of independent censoring:} \\
& \quad \alpha_i(t)dt \;\; E(I(\tilde{T}_i \ge t) | \mathcal{F}_{t-}) \\
&= \alpha_i(t)dt \;\; I(\tilde{T}_i \ge t) \; (\textit{the event } \{\tilde{T}_i \ge t\} \textit{ is known at } t-) \\
&= \alpha_i(t)dt \;\; Y_i(t)
\end{aligned}
$$

(the above shows in fact the equivalence in $(^*)$)

Left truncation, p. 4-5 in book

- *In a clinical study, the patients come under observation some time after the initiating event (i.e. the event defining $t = 0$)*
- *If time t is age, the individuals may be under observation from different ages*



Independent left truncation and right-censoring

The counting process results considered so far generalize almost immediately when **left-truncation** is included.

Under *left-truncation*, the observation for the ith individual is

$$(V_i, \tilde{T}_i, D_i)$$

where V_i is the *left-truncation* time (i.e., the time of entry) for the individual, and $\tilde{T}_i$ and D_i are as before.

We then have **independent left-truncation and right-censoring** provided that the counting process

$$N_i(t) = I\{\tilde{T}_i \leq t, D_i = 1\}$$

has intensity process given by $\lambda_i(t) = \alpha_i(t) Y_i(t)$, where the at risk indicator now takes the form

$$Y_i(t) = I\{V_i < t \leq \tilde{T}_i\}$$

Revisiting martingale theory

Recall the definition of intensity: $P(dN(t) = 1|\mathcal{F}_{t-}) = \lambda(t)dt$.

Since $dN(t)$ has possible values 0 and 1, this is equivalent to

$$E(dN(t)|\mathcal{F}_{t-}) = \lambda(t)dt$$

which implies

$$E(dN(t) - \lambda(t)dt|\mathcal{F}_{t-}) = 0 \text{ (since } \lambda(t) \text{ predictable)} \qquad (^*)$$

Thus if we introduce the process

$$M(t) = N(t) - \int_0^t \lambda(s)ds$$

then $M(t)$ is a martingale since

$$dM(t) = dN(t) - \lambda(t)dt$$

satisfies (*).

... revisiting martingale theory

Recall from previous slide that

$$M(t) = N(t) - \int_0^t \lambda(s)ds$$

is a (zero-mean) martingale. That is,

$$\begin{aligned} N(t) &= \int_0^t \lambda(s)ds \quad + \quad M(t) \\ &= \text{predictable increasing process} + \text{zero-mean martingale} \end{aligned}$$

which is exactly the unique Doob-Meyer decomposition.

Recall that $\Lambda(t) \equiv \int_0^t \lambda(s)ds$ is called the **compensator** of $N(t)$.

See Slides 5 p. 39++ where we also showed that $M(t) = N(t) - \lambda t$ is a martingale when $N(t)$ is a Poisson process.

Variation of counting processes

We now reinterpret concepts for general martingales to the case of counting process martingales.

Predictable variation process $\langle M \rangle (t)$

Recall the general definition

$$d \langle M \rangle (t) = Var(dM(t)|\mathcal{F}_{t-})$$

For the counting process case we have

$$dM(t) = dN(t) - \lambda(t)dt$$

so (see ABG p. 54)

$$\begin{aligned} d \langle M \rangle (t) &= Var(dM(t)|\mathcal{F}_{t-}) = Var(dN(t) - \lambda(t)dt|\mathcal{F}_{t-}) \\ &= Var(dN(t)|\mathcal{F}_{t-}) \end{aligned}$$

because $\lambda(t)dt$ is a 'constant' in this variance (why?) Hence, informally,

$$d \langle M \rangle (t) \approx \lambda(t)dt(1 - \lambda(t)dt) \approx \lambda(t)dt$$

and we conclude that $\langle M \rangle (t) = \int_0^t \lambda(s)ds = \Lambda(t)$.

Variation of counting processes

Recall from last slide, the **predictable** variation process is

$$\langle M \rangle (t) = \int_0^t \lambda(s) ds = \Lambda(t)$$

Optional variation process $[M](t)$

$$\begin{aligned} [M](t) &= \sum_{s \le t} (M(s) - M(s-))^2 \text{ (general)} \\ &= N(t) \text{ (counting process)} \end{aligned}$$

Recall also the general property

$$Var(M(t)) = E \langle M \rangle (t) = E[M](t)$$

which for counting processes becomes

$$Var(M(t)) = E(\Lambda(t)) = E(N(t))$$

We can recognize this for the Poisson process, where $N(t)$ is Poisson-distributed with expected value and variance λt.

Stochastic integral of general counting process

Consider a general counting process $N(t)$ and recall the decompositiom

$$M(t) = N(t) - \int_0^t \lambda(s)ds$$

Let further $T_1 < T_2 < \cdots$ be the jump times of $N(t)$.

Then a stochastic integral $\int_0^t H(s)dM(s)$ for a predictable process $H(s)$ can be handled as follows:

$$\begin{aligned} I(t) &= \int_0^t H(s)dM(s) \\ &= \int_0^t H(s)dN(s) - \int_0^t H(s)\lambda(s)ds \\ &= \sum_{T_j \le t} H(T_j) - \int_0^t H(s)\lambda(s)ds \end{aligned}$$

Variation processes of stochastic integrals of counting process martingales

Recall first *general facts*:

$$\left\langle \int H dM \right\rangle = \int H^2 d \langle M \rangle$$
$$\left[\int H dM \right] = \int H^2 d [M]$$

and the *special facts* for counting processes

$$\langle M \rangle (t) = \int_0^t \lambda(s) ds$$
$$[M](t) = N(t)$$

From this we get the formulas:

$$\left\langle \int H dM \right\rangle (t) = \int_0^t H^2(s)\, \lambda(s) ds$$
$$\left[\int H dM \right] (t) = \int_0^t H^2(s)\, dN(s)$$

The multiplicative intensity model

Assume that the counting process $N(t)$ has intensity function

$$\lambda(t) = \alpha(t)Y(t)$$

where $\alpha(t)$ is a *deterministic* function representing the risk (hazard rate) for each unit under study, and $Y(t)$ is the predictable process counting the number of units which are present immediately before t ("at risk") and hence may fail at time t.

We have the general expression

$$N(t) = \int_0^t \lambda(s)ds + M(t)$$

so we can write

$$dN(s) = \lambda(s)ds + dM(s) = \alpha(s)Y(s) + dM(s) \quad (^*)$$

Our goal is to estimate $\alpha(t)$, but it turns out to be easier to estimate its integral $A(t) = \int_0^t \alpha(s)ds$, which we will do.

Assume for simplicity that we know that $Y(t) > 0$ for all t. Then - dividing through the equation $(^*)$ by $Y(t)$ we get

The multiplicative intensity model (cont.)

$$\frac{1}{Y(s)}dN(s) = \alpha(s)ds + \frac{1}{Y(s)}dM(s)$$

Integrating we get that

$$\int_0^t \frac{1}{Y(s)}dN(s) = A(t) + \int_0^t \frac{1}{Y(s)}dM(s)$$

The rightmost expression is a mean zero integral (since $Y(s)$ is a predictable process), so this suggests the (Nelson-Aalen) estimator

$$\hat{A}(t) = \int_0^t \frac{1}{Y(s)}dN(s) = \sum_{T_j \leq t} \frac{1}{Y(T_j)}$$

Properties of the estimator

We have

$$\hat{A}(t) - A(t) = \int_0^t \frac{1}{Y(s)} dM(s)$$

and hence $\hat{A}(t)$ is an unbiased estimator (since the right hand side is a mean zero martingale). Further (see earlier page 18)

$$\left\langle \hat{A} - A \right\rangle (t) = \int_0^t \frac{1}{Y^2(s)} \alpha(s) Y(s) ds = \int_0^t \frac{1}{Y(s)} \alpha(s) ds$$

$$\left[\hat{A} - A \right] (t) = \int_0^t \frac{1}{Y^2(s)} dN(s) = \sum_{T_j \leq t} \frac{1}{Y^2(T_j)} \tag{1}$$

We also know from our earlier formulas that

$$Var(\hat{A}(t)) = \mathrm{Var}(\hat{A}(t) - A(t)) = E\left(\left[\hat{A} - A \right] (t) \right)$$

so that (1) is an unbiased estimator of $Var(\hat{A}(t))$

Asymptotic distribution of Nelson-Aalen estimator

Suppose that our data were based on observation of a large number n units. We shall let n tend to infinity in the relation

$$\sqrt{n}(\hat{A}(t) - A(t)) = \int_0^t \sqrt{n}\frac{1}{Y(s)}dM(s)$$

and hence we need to look at the limiting behaviour of the mean zero martingale on the right hand side.

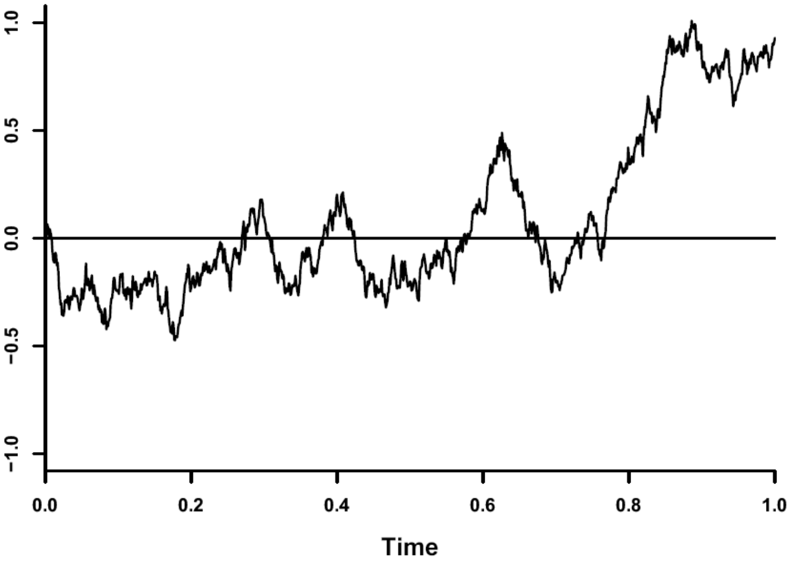
This brings us to the need for an asymptotic theory for martingales.

Wiener process

A Wiener process (Brownian motion) with variance parameter 1 is a stochastic process $W(t)$ with values in the real numbers satisfying

1. $W(0) = 0$
2. For any $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \cdots \leq s_n \leq t_n$, the random variables $W(t_1) - W(s_1), \ldots, W(t_n) - W(s_n)$ are independent.
3. For any $s < t$, the random variables $W(t) - W(s)$ are normal with expected value 0 and variance $(t - s)$
4. The paths are continuous.

Wiener process



Gaussian martingales

Let $V(t)$ be a strictly increasing continuous function with $V(0) = 0$ and consider the stochastic process

$$U(t) = W(V(t))$$

(i.e. a time transformation of $W(t)$.)

The process $U(t)$ is a Gaussian process which is moreover

- a mean zero martingale
- has predictable variation process $\langle U \rangle (t) = V(t)$

(exercise 2.12 in book).

U is called a Gaussian martingale.

Rebolledo's martingale convergence theorem

Let $\tilde{M}^{(n)}(t)$ be a sequence of mean zero martingales for $t \in [0, \tau]$.

ABG gives a precise formulation of conditions under which $\tilde{M}^{(n)}(t)$ converges in distribution to a Gaussian martingale $U(t)$ as defined on the previous page.

What is needed is that as $n \to \infty$,

(i) $\left\langle \tilde{M}^{(n)}(t) \right\rangle \to V(t)$ in probability for all $t \in [0, \tau]$ as $n \to \infty$

(ii) The sizes of the jumps of $\tilde{M}^{(n)}(t)$ go to zero (in probability)

Application to counting process martingales

We will consider martingales of the form

$$\tilde{M}^{(n)}(t) = \int_0^t H^{(n)}(s)dM^{(n)}(s)$$

where $H^{(n)}(s)$ is a predictable process and

$$M^{(n)}(t) = N^{(n)}(t) - \int_0^t \lambda^{(n)}(s)ds$$

is a counting process martingale.

Assume that we can write $V(t) = \int_0^t v(s)ds$. Then sufficient conditions for Rebolledo's theorem are

(i) $(H^{(n)}(s))^2\lambda^{(n)}(s) \to v(s) > 0$ for all $s \in [0, \tau]$

(ii) $H^{(n)}(s) \to 0$ for all $s \in [0, \tau]$

Nelson-Aalen example

Recall that

$$\sqrt{n}(\hat{A}(t) - A(t)) = \int_0^t \frac{\sqrt{n}}{Y(s)} dM(s)$$

so that we have

$$H^{(n)}(t) = \frac{\sqrt{n}}{Y(t)}$$

Assume that there is a deterministic positive function $y(t)$ such that $Y(t)/n \to y(t) > 0$ in probability. Then the two sufficient conditions for Rebolledo's theorem are satisfied:

$$(H^{(n)}(s))^2 \lambda^{(n)}(s) = \frac{n}{Y^2(s)} \cdot \alpha(s) Y(s) = \frac{\alpha(s)}{Y(s)/n} \to \frac{\alpha(s)}{y(s)} \equiv v(s)$$

$$H^{(n)}(s) = \frac{\sqrt{n}}{Y(s)} = \frac{1/\sqrt{n}}{Y(s)/n} \to 0$$

Nelson-Aalen example (cont.)

Conclusion:

$$\sqrt{n}(\hat{A}(t) - A(t)) = \int_0^t \frac{\sqrt{n}}{Y(s)} dM(s)$$

converges in distribution to the mean zero Gaussian martingale $U(t) = W(V(t))$ with predictable variation process

$$V(t) = \int_0^t v(s)ds = \int_0^t \frac{\alpha(s)}{y(s)} ds$$

In particular, for each t is $\hat{A}(t)$ asymptotically normally distributed, and

$$\mathrm{Var}(\hat{A}(t)) \approx \frac{1}{n} \cdot \int_0^t \frac{\alpha(s)}{Y(s)/n} ds = \int_0^t \frac{\alpha(s)}{Y(s)} ds = \left\langle \hat{A} - A \right\rangle (t)$$