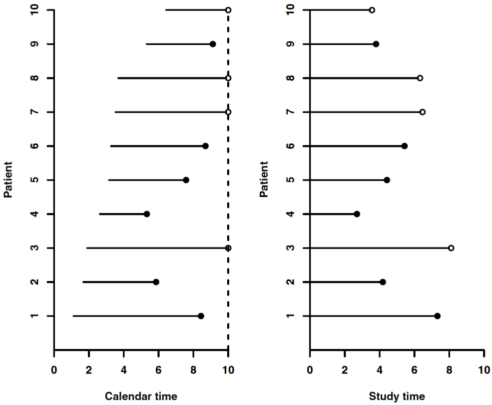
Section 10

Lecture 5

Plan for today

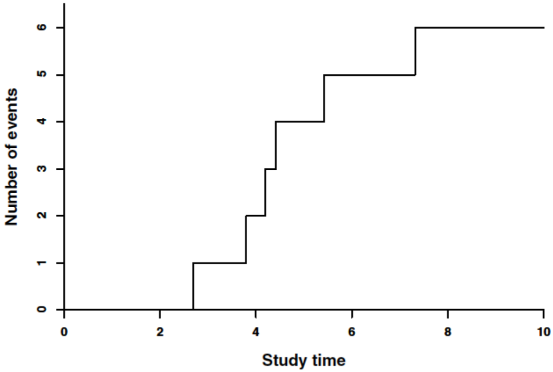
- Counting processes
- Independent censoring
  - Formal conditions
  - A proof
  - Conditions that can be evaluated in graphs
- Multiplicative intensity model
- The Nelson-Aalen estimator

Example



7.32, 4.19, 8.11, 2.70, 4.42, 5.43, 6.46, 6.32, 3.80, 3.50.

Example: Counting process description



7.32, 4.19, 8.11, 2.70, 4.42, 5.43, 6.46, 6.32, 3.80, 3.50.

What is a counting process?

Definition

A counting process is a right-continuous stochastic process $\{N(t); t \geq 0\}$ with jumps of $+1$. It satisfies

- $N(0) = 0$, $N(t) \geq 0$, $t \geq 0$,
- $N(t)$ is an integer,
- if $s \leq t$ then $N(s) \leq N(t)$.

Discrete state space, but right-continuous sample paths.

Illustration of counting process

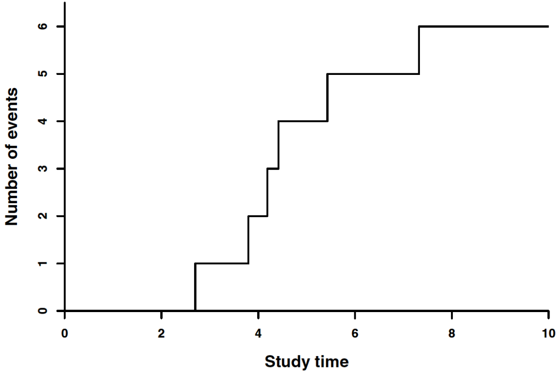


Illustration of a counting process

The point is: analyses of classical medical studies often lead to outcomes that can be represented as counting processes.

Using the Doob-Meyer decomposition to define intensities

- The Doob-Meyer decomposition ensures that there exists a unique predictable process $\Lambda(t)$ such that $M(t) = N(t) - \Lambda(t)$ is a mean zero martingale.
- Suppose that $\Lambda(t)$ is absolutely continuous. [15] Then, there exists a predictable process $\lambda(t)$ such that

$$\Lambda(t) = \int_0^t \lambda(s) ds.$$

- $\lambda(t)$ is the **intensity**.
- $\Lambda(t)$ is the cumulative intensity.

[15] We will assume this throughout, unless otherwise stated.

Some observations...

- $[M](t) = N(t), \ \forall t > 0$, when N is a counting process and M is a martingale given by the Doob-Meyer decomposition,

  $$M(t) = N(t) - \Lambda(t),$$

  because only the jump remains in the limit

  $$[M](t) = \lim_{n\to\infty} \sum_{k=1}^{n} (\Delta M_k)^2.$$

- $\langle M \rangle(t) = \int_0^t \lambda(s) ds$.
  Argument in the next slide.
  It follows that $\Lambda(t)$ is a compensator of $N(t)$ and $M(t)^2$. Useful, remember that $\text{Var}(M(t)) = \mathbb{E}\{M(t)^2\}) = \mathbb{E}\{\langle M \rangle(t)\} = \mathbb{E}\{[M](t)\}$.
- This is similar to a Poisson process (a homework question)!

Informal argument for $\langle M\rangle(t) = \int_0^t \lambda(s)ds$.

Let $d\langle M\rangle(t) = VAR(dM(t) \mid \mathcal{F}_{t-})$ be the increment of the predictable variation in a small interval $[t, t + dt)$.
Consider the following heuristic argument

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

Remember that $dN(t) \in \{0, 1\}$ and thus (informally)

$$\lambda(t)dt = P(dN(t) = 1 \mid \mathcal{F}_{t-}) = \mathbb{E}(dN(t) \mid \mathcal{F}_{t-}),$$

and

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

From survival times to counting processes

Let us explicitly consider the relation between survival times and counting processes.

- Consider n individuals with survival times $T_1, T_2, \ldots, T_n$.[16]
- Suppose that these survival times are *independent* and that T_i is distributed according to hazard $\alpha_i(t)$.
- Define the individual basic (uncensored) process $N_i^c(t) = I(T_i \leq t)$.[17]
- Define the filtration $\{\mathcal{F}_t^c\}$ is an increasing family of σ algebras generated by $N_i^c(t)$.[18]

[16] When not otherwise stated, we will assume that T_i is absolutely continuous and thus the events *do not* happen at the same time w.p.1, $T_i \neq T_j \forall i, j$.

[17] Here superscript "c" denotes complete, to highlight that this is the count of the event process, that might be unobserved due to censoring.

[18] the *generated* filtration associated to a stochastic process is a filtration which records the "past behaviour" of the process at each time.

From survival times to counting processes (informally)

- Let, $dN_i^c(t)$ denote the number of jumps of the process in a small interval $[t, t + dt)$, such that only a single event can occur in the interval. Then, heuristically,

$$P(dN_i^c(t) = 1 \mid \mathcal{F}_{t-}^c) = P(t + dt > T_i \geq t \mid \mathcal{F}^c_{t-}) = \begin{cases} \alpha_i(t)dt, \, T_i \geq t, \\ 0, \, T_i < t. \end{cases}$$

- The intensity process $\lambda_i^c(t)$ is

$$\begin{aligned} \lambda_i^c(t)dt &= P(dN_i^c(t) = 1 \mid \mathcal{F}_{t-}^c). \\ &= \mathbb{E}(dN_i^c(t) \mid \mathcal{F}_{t-}^c) \text{ bc. } dN_i^c(t) \text{ is binary.} \end{aligned}$$

Note that:

We can write $\lambda_i^c(t)$ for $i = 1, 2, \dots$ on the multiplicative form

$$\lambda_i^c(t) = \alpha_i(t) I(T_i \geq t),$$

where $\alpha_i(t)$ is the hazard rate.
If T_1, $T_2, \dots, T_n$ are i.i.d. we can indeed write

$$\lambda_i^c(t) = \alpha(t) I(T_i \geq t).$$

Counting process for a survival function

Suppose we have survival times $T_1, T_2, \ldots, T_n$ corresponding to the survival times of n independent individuals.

- Define the aggregated process $N^c(t) = \sum_{i=1}^n N_i^c(t)$.
  that counts the number of events in the population, e.g. deaths in a medical study.
- For i.i.d. individuals we have

$$\lambda^c(t)dt = \alpha(t) \sum_{i=1}^{n} I(T_i \geq t)dt \text{ when } \alpha_i(t) = \alpha(t).$$

Illustration: Aggregated survival

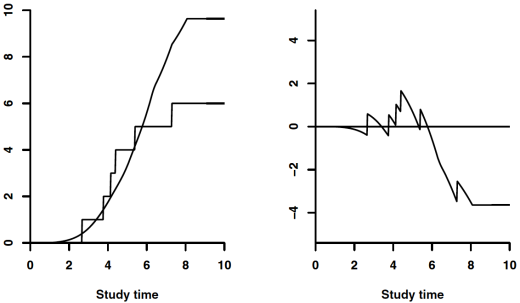


Fig. 1.14 *Illustration of a counting process and its cumulative intensity process (left panel) and the corresponding martingale (right panel).*

Right censoring

- We follow subjects over time and observe $(\tilde{T}_i, D_i)$,

$$\tilde{T}_i = T_i \text{ if } D_i = 1,$$
$$\tilde{T}_i < T_i \text{ if } D_i = 0.$$

- Indeed, $\tilde{T}_i = T_i \wedge T_i^* = \min(T_i, T_i^*)$, where T_i^* is called the censoring time.
- We define the right censoring process
  $C_i(t) = I(t > T_i^*) = 1 - I(t \leq T_i^*)$.
  This process is **left** continuous.
- Let $Z_i(t) = I(t \leq \tilde{T}_i)$.
  Process denoting "no event yet"

Before we introduce independent censoring

To be clear, let L_0 be a set of baseline covariates and let's write out some explicit examples of filtrations:

$\mathcal{N}_t = \sigma(N(u); 0 \leq u \leq t)$ (sometimes called the self-exciting filtration)
$\mathcal{N}_t^c = \sigma(N^c(u); 0 \leq u \leq t)$ (another self-exciting filtration)
$\mathcal{F}_0^c = \sigma(L_0, A)$ and $\mathcal{F}_t^c = \sigma(L_0, N^c(u); 0 \leq u \leq t)$.
$\mathcal{G}_t = \sigma(A, L_0, N^c(u), C(u); 0 \leq u \leq t)$ so $\{\mathcal{G}_t^c\} \supseteq \{\mathcal{F}_t^c\}$.
$\mathcal{F}_t = \sigma(A, L_0, N(u), Z(u); 0 \leq u \leq t)$.

Some important things to remember

- A counting process is a right-continuous stochastic process $\{N(t); t \geq 0\}$ with jumps of $+1$.
- The Doob-Meyer decomposition ensures that there exists a unique predictable process $\Lambda(t)$ such that $M(t) = N(t) - \Lambda(t)$ is a mean zero martingale.
- Martingales are generalizations of random errors, and a lot of nice theory is developed for martingales.
- Useful, remember that $\text{Var}(M(t)) = \mathbb{E}\{M(t)^2\}) = \mathbb{E}\langle M\rangle(t) = \mathbb{E}\{[M](t)\}$.
- We follow subjects over time and observe $(\tilde{T}_i, D_i)$,

$$\tilde{T}_i = T_i \ \text{ if } \ D_i = 1,$$
$$\tilde{T}_i < T_i \ \text{ if } \ D_i = 0.$$

- Indeed, $\tilde{T}_i = T_i \wedge T_i^* = \min(T_i, T_i^*)$, where T_i^* is the censoring time.

Independent censoring (technical definition)

Definition (Independent censoring, Andersen et al)

Let N^c be the basic (uncensored) counting process with compensator Λ^c (i.e. cumulative intensity) with respect to a given filtration $\{\mathcal{F}_t^c\}$. Let C be a right-censoring process which is predictable with respect to a filtration $\{\mathcal{G}_t\} \supseteq \{\mathcal{F}_t^c\}$. Then we call the right-censoring of N generated by C **independent** if the compensator of N^c with respect to $\mathcal{G}_t$ is also Λ^c.

Intuition (i): keep the risk sets (i.e. those who are alive and not censored) representative for the whole population.
Intuition (ii): Knowledge of the censoring times does not alter the intensity process for N.
This definition is quite abstract. Perhaps unnecessarily abstract. I think a better definition is given by thinking causally. In the next slide, we do this.

A better way of encoding and grasping independent censoring is to consider causal graphs

- The heuristic idea is to ensure that the censoring variable is *independent* of the future outcome variables, given the measured past (the filtration).
- We can state this as an independence condition with respect to the outcome of interest (in a discrete time setting).
- The history of a random variable through k is denoted by an overbar, e.g. $\overline{L}_k \equiv (L_0, \dots, L_k)$, and future events are denoted by underbars, e.g. $\underline{N}_k \equiv (N_k, \dots, N_K)$.

A setting where we have access to baseline covariates L_0 and treatment A, it is sufficient to evaluate:

$$\underline{N}_k^c \perp\!\!\!\perp C_k \mid L_0, N_{k-1} = C_{k-1} = 0, A \tag{8}$$

where $L_0, \overline{N}_k, \overline{C}_k$ can be interpreted as a discrete filtration, i.e. everything that is measured in the past.
This assumption requires that at each follow-up time, given the measured past, censoring is independent of future counterfactual outcomes had censoring been eliminated.

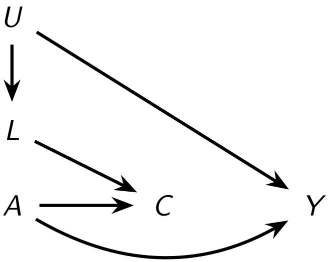
Time-varying version

A setting where we have access to time-varying covariates L_k.

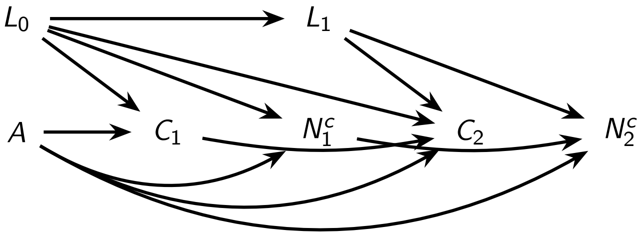
$$\underline{N}_k^c \perp\!\!\!\perp C_k \mid \overline{L}_k, N_{k-1} = C_{k-1} = 0, A \tag{9}$$

where $\overline{L}_k = \overline{l}_k, L_k = C_k$ can be interpreted as a discrete filtration, i.e. everything that is measured in the past.

Our graph from the HIV example



Time-varying version (U omitted to avoid clutter)



Here, we can use d-separation to read off that

$$\underline{N}_1^c \perp\!\!\!\perp C_1 | L_0, A, \tag{10}$$

and

$$N_2^c \perp\!\!\!\perp C_2 | \overline{L}_1, N_1^c = C_1 = 0, A \tag{11}$$

How to generate random censoring times?

Simple example

- Draw n independent variables $T_i^* \sim \mathcal{P}_c$ where $\mathcal{P}_c$ has support $[0, \tau)$.
- For each $i = 1, \ldots, n$, draw $T_i \sim \mathcal{P}_n$ independently of T_i^* where P_n has support $[0, \tau)$.
- Set $\tilde{T}_i = \min(T_i^*, T_i)$ and set $D_i = I(T_i \leq T_i^*)$.

Argument for independent censoring

We will use the innovation theorem.

Theorem (Innovation theorem)

An intensity $\lambda_i^{\mathcal{F}''}(t)$ *with respect to a filtration* $\{\mathcal{F}_t''\}$ *such that* $\{\mathcal{F}_t'\} \supseteq \{\mathcal{F}_t''\}$*, satisfies*

$$\lambda_i^{\mathcal{F}''}(t) = \mathbb{E}(\lambda_i^{\mathcal{F}'}(t) \mid \mathcal{F}_{t-}'').$$

Proof.

We use iterative expectations for small dt (informally),

$$\lambda_i^{\mathcal{F}''}(t)dt = \mathbb{E}(dN_i(t) \mid \mathcal{F}_{t-}'') = \mathbb{E}\{\mathbb{E}(dN_i(t) \mid \mathcal{F}_{t-}') \mid \mathcal{F}_{t-}''\},$$

and the result follows because

$$\lambda_i^{\mathcal{F}'}(t)dt = \mathbb{E}(dN_i(t) \mid \mathcal{F}_{t-}').$$

□

Independent censoring allows us to write....

Theorem (identification under independent censoring)

Under independent censoring, the intensity of the right-censored counting process N_i can be written as

$$\lambda_i(t)dt = Z_i(t)\alpha_i(t)dt$$

where $Z_i(t) = I(t \leq \tilde{T}_i)$ and α_i is the hazard of the "complete" counting process

$$\lambda_i^c(t)dt = Z_i^c(t)\alpha_i(t)dt$$

where $Z_i^c(t) = I(t \leq T_i)$.

Thus, importantly, we can identify α_i from the **censored** data under independent censoring.

Proof independent censoring

Proof.

As before, define $C_i(t) = I(t > T_i^*)$, which is left continuous and $\mathcal{F}_t$ predictable.
Define $Z_i(t) = I(t \leq \tilde{T}_i)$.

$$
\begin{aligned}
\lambda_i^{\mathcal{F}}(t)dt &= \mathbb{E}(dN_i(t) \mid \mathcal{F}_{t-}) = \mathbb{E}(dN^c(t)[1 - C_i(t)] \mid \mathcal{F}_{t-}) \\
&= \mathbb{E}\{\mathbb{E}(dN^c(t)[1 - C_i(t)] \mid \mathcal{G}_{t-}) \mid \mathcal{F}_{t-}\} \text{ (iterative expectation)} \\
&= \mathbb{E}\{[1 - C_i(t)]\mathbb{E}(dN^c(t) \mid \mathcal{G}_{t-}) \mid \mathcal{F}_{t-}\} \text{ (C_i is predictable)} \\
&= \mathbb{E}\{[1 - C_i(t)]\lambda_i^{\mathcal{G}}(t)dt \mid \mathcal{F}_{t-}\} \text{ (innovation theorem)} \\
&= \mathbb{E}\{I(t \leq T_i^*)I(t \leq T_i)\alpha_i(t)dt \mid \mathcal{F}_{t-}\} \text{ (independent censoring)} \\
&= \mathbb{E}\{Z_i(t)\alpha_i(t)dt \mid \mathcal{F}_{t-}\} \\
&= Z_i(t)\mathbb{E}\{\alpha_i(t)dt \mid \mathcal{F}_{t-}\} \text{ (Z_i is $\mathcal{F}_t$ predictable)} \\
&= Z_i(t)\alpha_i(t)dt \text{ (assuming α_i is $\mathcal{F}_t$ predictable).}
\end{aligned}
$$

Section 11

Estimation

Multiplicative intensity model

Definition (Multiplicative intensity model)

The multiplicative intensity model is the class of statistical models that has an intensity process $\lambda(t)$ of a counting process N wrt. a filtration $\{\mathcal{F}_t\}$ where

$$\lambda(t) = \alpha(t)Z(t),$$

such that $\alpha(t)$ is a non-negative deterministic function and $Z(t)$ is a (left-continuous and adapted) predictable process that does not depend on $\alpha(t)$.

Note that:

- $Z(t)$ is observable
- Remember that $Z(t) = \sum_{i=1}^{n} I(\tilde{T}_i \geq t)$ in our example with censored survival times with $\alpha_i(t) = \alpha(t)$.
- $Z_i(t) = I(T_i \geq t)$ in the *uncensored* survival examples before.

The Nelson-Aalen estimator

Let $T_1 < T_2 < \dots$ be jump times of a counting process $N(t)$ with an intensity that satisfies the multiplicative intensity model.

Definition (Nelson-Aalen Estimator)

The Nelson-Aalen estimator of $H(t) = \int_0^t \alpha(s) ds$ is
$\hat{H}(t) = \sum_{T_j \leq t} \frac{1}{Z(T_j)} \equiv \sum_{T_j \leq t} \Delta \hat{H}(T_j)$,

where $Z(t)$ is an "at risk" process, where $Z(t) \geq 0$.
As before, we will let $Z(t) = \sum_{i=1}^n I(\tilde{T}_i \geq t)$.
That is, the estimator is a weighted sum over the jump times of N.
Thus, the Nelson-Aalen estimator, $\hat{H}(t) = \sum_{T_j \leq t} \frac{1}{Z(T_j)}$, is a counting process integral, where $\frac{1}{G(t)} = Z(t) = \sum_{i=1}^n I(\tilde{T}_i \geq t)$.

- This is a non-parametric estimator; not imposed structure on $\alpha(t)$ or $H(t)$.

Example of multiplicative intensity model:

- We have already seen that: $\lambda_i = \alpha_i(t) I(\tilde{T}_i \geq t)$ wrt. $\mathcal{F}_t^c$ under independent censoring
- Suppose that $\lambda_i = \alpha(t) I(\tilde{T}_i \geq t)$, then
  $\lambda(t) = \sum_{i=1}^n \lambda_i(t) = \alpha(t) Z(t)$,
  where $\alpha(t)$ is non-negative and $Z(t) = \sum_{i=1}^n Z_i(t)$ is the number at risk just before t.

  This holds when we have i.i.d. individuals such that the survival times T_i is distributed with hazard $\alpha(t)$.

Transformation of a martingale

Let $G = \{G_0, G_1, G_2, \dots\}$ be a predictable stochastic process and $M = \{M_0, M_1, M_2, \dots\}$ be a martingale wrt $\{\mathcal{F}_n\}$.
Define $Z = \{Z_0, Z_1, Z_2, \dots\}$ by

$$Z_n = G_0 M_0 + G_1(M_1 - M_0) + \cdots + G_n(M_n - M_{n-1}).$$

If $M_0 = 0$, then Z is a mean zero martingale,

$$\mathbb{E}(Z_n - Z_{n-1} \mid \mathcal{F}_{n-1}) = \mathbb{E}(G_n(M_n - M_{n-1}) \mid \mathcal{F}_{n-1}) = G_n \mathbb{E}(M_n - M_{n-1} | \mathcal{F}_{n-1}) = 0.$$

The process Z is denoted the transformation of M by G and it is written $Z = G \bullet M$. There is a close connection to stochastic integration here, as we will see. Stochastic integration is integration of one stochastic process with respect to another stochastic process. The importance of stochastic integrals will be clear when we study estimators.

Stochastic integrals for counting processes

- Let $G = \{G(t) : t \in [0, \tau]\}$ be a predictable stochastic process and $M = \{M(t) : t \in [0, \tau]\}$ be a mean zero martingale wrt $\{\mathcal{F}_t\}$.
- Consider the stochastic integral for a *counting process Martingale M*,

$$
\begin{aligned}
I(t) &= \int_0^t G(s)dM(s) \\
&= \lim_{n \to \infty} \sum_{k=1}^{n} G_k \Delta M_k,
\end{aligned}
$$

  where $[0, t]$ is partitioned into n subintervals of length t/n and $G_k = G((k-1)t/n)$ and $\Delta M_k = M(kt/n) - M((k-1)t/n)$.[20]
  This is a Stieltjes integral.
- Importantly, $I(t)$ is a mean zero martingale wrt $\{\mathcal{F}_t\}$.
  This is the limit of discrete time transformations (Slide 151)

[20] In general, the limiting distribution is not valid and we must introduce Itô integrals.

Stochastic integrals for counting processes

However, from the Doob-Meyer decomposition we see that

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

- Here, $\int_0^t G(s)dN(s) = \sum_{T_j \le t} G(T_j)$ is denoted the counting process integral of G.