Biostatistics

Mats J. Stensrud

Spring 2021

Section 1

Structure of the course

Structure

- 90 minutes lectures every Tuesday 10h15.[1]
- Moodle is our platform.
  - Announcements
  - Links to relevant literature
  - Link to Piazza (Password: SURVIVAL).
- Slides, video recordings and problem sheets will be uploaded every Tuesday.

[1] However, if people do not meet up, I will pre-record the lectures.

Exam

- Written exam.
- Two graded homeworks.

Features of the course

- This is a *statistics* course.
- We will study theory and methods that is relevant to solve common practical problems.
- The course will contain proofs,
  but *all* the results we are using will *not* be shown.
  That said, I will strive to motivate all the results.
- I will also spend time on discussing the interpretation of the results.

Goal

After the course, you should be able to:

- Understand mathematical and statistical theory for event history analysis and longitudinal data analysis.
- Furthermore, understand the concepts and ideas that this theory expresses.
- Apply these methods to data (there are ubiquitous applications!).
- Critically evaluate how these methods are used in practice.
- Build on the material in this course to derive new results yourself.

Outline of the course

- **Time-to-event outcomes** ("survival analysis")
- Longitudinal data analysis
- Research synthesis

Section 2

Time-to-events and survival analysis

The Moderna vaccine

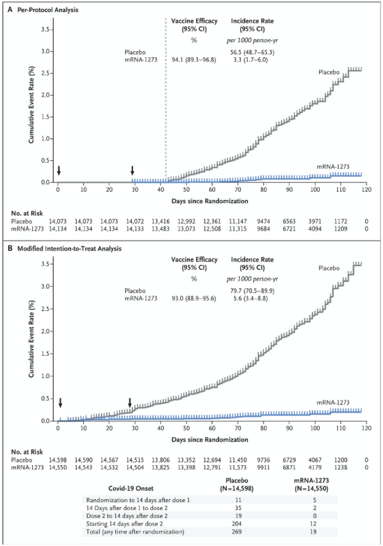


Figure 1: Survival analysis is e.g. used to present results from vaccine trials.

Time to events are all over the place

- Time from birth to death.
- Time from birth to cancer diagnosis.
- Time from disease onset to death.
- Time from entry to a study to relapse.
- Time from marriage to divorce.
- Time from production until a machine is broken.
- Time from origin of the coronavirus until a stock (marked) crashes.

NEWS FEATURE

THE TOP 100 PAPERS

Nature *explores the most-cited research of all time.*

STATISTICS

Although the top-100 list has a rich seam of papers on statistics, says Stephen Stigler, a statistician at the University of Chicago in Illinois and an expert on the history of the field, "these papers are not at all those that have been most important to us statisticians". Rather, they are the ones that have proved to be most useful to the vastly larger population of practising scientists.

Much of this crossover success stems from the ever-expanding stream of data coming out of biomedical labs. For example, the most frequently cited statistics paper (number 11) is a 1958 publication[15] by US statisticians Edward Kaplan and Paul Meier that helps researchers to find survival patterns for a population, such as participants in clinical trials. That introduced what is now known as the Kaplan–Meier estimate. The second (number 24) was British statistician David Cox's 1972 paper[16] that expanded these survival analyses to include factors such as gender and age.

Figure 2: The two most cited statistics papers concern survival analysis

Some common questions

- What is survival under treatment A vs B?
- What is the duration of a certain component in the machine?
- How long does it take before a stock marked crashes?

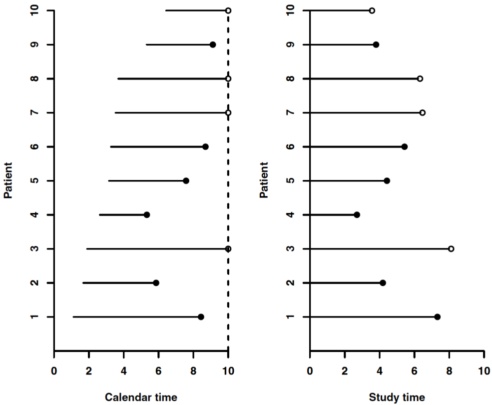
`PS: These questions are very often about causal effects....`

An overview of the data structure 2

- We follow units of over time;
  humans, animals, engines, etc.
- The events of interest may be **the time to** deaths, cancer diagnoses, divorces, child births, engine failures, etc.
- We often stop the study before everyone has experienced the event of interest.

Censored survival times (illustration)

Consider 10 patients with newly diagnosed cancer. Let $T \in (0, \tau]$ be a survival time.



7.32, 4.19, 8.11, 2.70, 4.42, 5.43, 6.46, 6.32, 3.80, 3.50.
How do you estimate $\mathbb{E}(T)$, that is, the mean survival?

Why not use "standard methods"?

- We have incomplete observations.
- Instead of observing T_i we observe $(\tilde{T}_i, D_i)$,

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

  where D_i is a censoring indicator.

  We want to use our information on $\tilde{T}_i$ to make inference on T.
- There is a strong link to causal inference and "what if" questions: What would happen if we observed T_i instead of $\tilde{T}_i$.
- We must make assumptions about the censoring, similarly to assumptions in causal inference.

Let's start with a single outcome process

Assume $T > 0$ is an absolutely continuous random variable.

Definition (Survival function)

The survival function is $S(t) = P(T > t)$, that is, the probability that the survival time T exceeds t.

Definition (Hazard rate)

The hazard rate $\alpha(t) = \lim_{dt \to 0} \frac{1}{dt} P(t + dt > T > t \mid T \geq t)$ is the rate of events per unit of time.
That is,.

Informally, $\alpha(t)dt = P(t + dt > T > t \mid T \geq t)$ is the probability that the event will happen between time t and time $t + dt$ given that it has not happened earlier.[2]

[2] PS: We are going to extend this to multiple events later.

Cumulative hazard and some relations

Define the cumulative hazard,

$$A(t) = \int_0^t \alpha(s)ds.$$

Then,

$$A'(t) = \alpha(t) = \lim_{dt \to 0} \frac{1}{dt} \frac{S(t) - S(t+dt)}{S(t)} = -\frac{S'(t)}{S(t)} = \frac{f(t)}{S(t)}.$$

By integration

$$\int_0^t \alpha(s)ds = -\log\{S(t)\},$$

and thus

$$S(t) = \exp\{-\int_0^t \alpha(s)ds\}.$$

$\alpha(t)$ completely determines the distribution of survival times T.

Illustration of hazards and survival functions

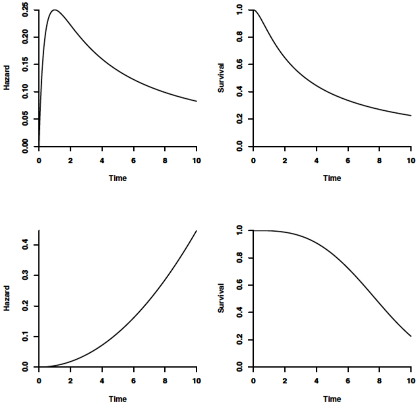
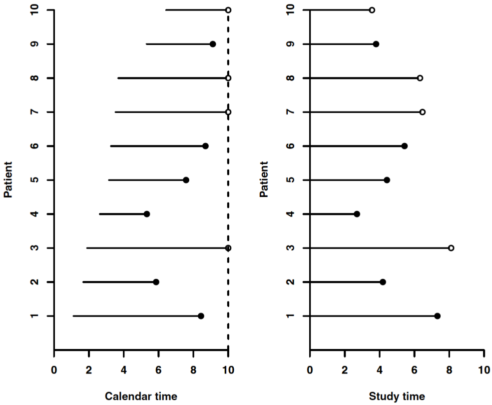


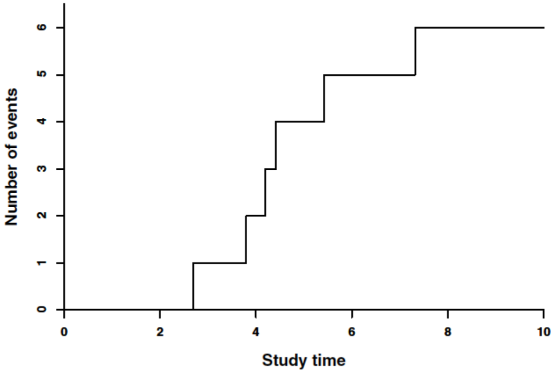
Fig. 1.2 *Illustrating hazard rates and survival curves. The hazard rates on the left correspond to the survival curves on the right.*

Example



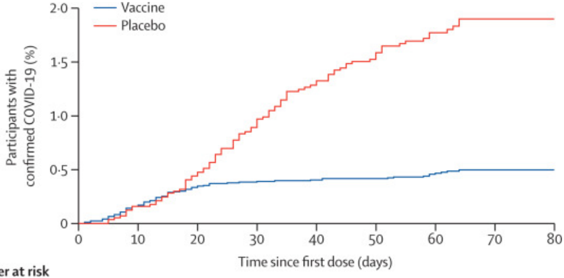
7.32, 4.19, 8.11, 2.70, 4.42, 5.43, 6.46, 6.32, 3.80, 3.50.

Example: Counting process description



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

Before we turn to more abstract things: Sputnik



Number at risk (number of COVID-19 cases)								
Vaccine	16 427 (0)	15 338 (35)	15 717 (61)	14 683 (66)	10 970 (70)	6686 (71)	3314 (77)	398 (79)
Placebo	5435 (0)	5121 (10)	5046 (30)	4895 (54)	3662 (71)	2223 (87)	1106 (92)	133 (96)

The Sputnik Vaccine.

Section 3

Processes

Many of you are familiar with stochastic processes

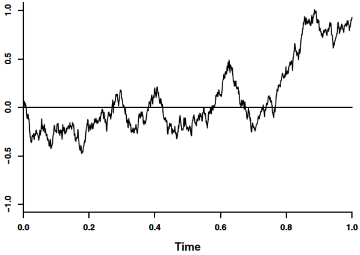
Here I will review *basic* concepts and results on stochastic processes.
I will give definitions and proceed at a "working technical" level.
We will focus on counting processes and martingales.
There are rigorous courses on stochastic processes at EPFL, such as:

- MATH-330 Martingales et mouvement brownien (Prof. Aru).
- MATH-332 Stochastic processes (Prof. Mountford).

Stochastic process

- A stochastic process is a time-indexed collection of random variables, say, $\{X(t) : t \in [0, \tau]\}$.
- Consider a probability space $(\Omega, \mathcal{F}, P)$.
- A filtration $\{\mathcal{F}_t\}_{t>0}$ is an increasing right-continuous family of sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$ whenever $s \leq t$.
  Think about the filtration as representing the **past**, that is, the history.
- We denote $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t>0}, P)$ a filtered probability space.

Example



Wiener process (Brownian motion)

Definition (Wiener process)

The $W = \{W(t) : t \in [0, \tau]\}$ is a process satisfying

- $W(0) = 0$,
- independent increments, that is, $W(t + u) - W(t)$ $u \geq 0$ are independent of $W(s)$, for all $s \leq t$,
- Gaussian increments, that is, $W(t + u) - W(t) \sim \mathcal{N}(0, u)$,
- continuous sample paths, that is, $W(t)$ is continuous t.

Adapted process

Definition (Adapted process)

A stochastic process $X = \{X(t); t \in [0, \tau]\}$ is adapted to $\{\mathcal{F}_t\}$ if $X(t)$ is $\mathcal{F}_t$ measurable for each t.

`Intuitively, the value of` $X(t)$ `is known at` t.

PS: We will also consider the stronger notion of a *predictable* processes. We omit a formal definition of *predictable* but state the sufficient conditions that a process $X = \{X(t); t \in [0, \tau]\}$ is predictable if

- X is adapted to $\{\mathcal{F}_t\}$, and
- the sample paths of X are left-continuous.[3]

`Intuitively, the value of` $X(t)$ `is known just before` t.

[3] A sample path is a realization of X, which is a function of t.

Martingale

Definition (Martingale)

A stochastic process $M = \{M(t); t \in [0, \tau]\}$ is a martingale relative to $\{\mathcal{F}_t\}$ if X is adapted to $\{\mathcal{F}_t\}$ and $\mathbb{E}(M(t) \mid \mathcal{F}_s) = M(s)$ for all $t > s$.

Informally, $\mathbb{E}(dM(t) \mid \mathcal{F}_{t-}) = 0$, where $\mathcal{F}_{t-}$ is the filtration just before t.
$\mathcal{F}_{t-}$ is the smallest σ algebra containing all $\mathcal{F}_s$, $s < t$.

Martingale intuition

Definition (Discrete martingale)

Let $M = \{M_0, M_1, M_2, \dots\}$ be a *discrete* stochastic process adapted to $\{\mathcal{F}_n\}$.
The discrete process M is a martingale if

$$\mathbb{E}(M_n \mid \mathcal{F}_{n-1}) = M_{n-1}$$

.

- Heuristic: Think about the Martingale as cumulative noise, similar to random errors in "standard" statistical models.

Some features of (discrete) Martingales

- The definition is equivalent to saying that $\mathbb{E}(M_n \mid \mathcal{F}_m) = M_m$ for $m < n$.
- Suppose $M_0 = 0$. Then $\mathbb{E}(M_n) = 0$
  because $\mathbb{E}(M_n) = \mathbb{E}(\mathbb{E}(M_n \mid \mathcal{F}_0)) = \mathbb{E}(M_0) = 0$.
- It also follows that

$$\text{Cov}(M_m, M_n - M_m) = 0, \forall n > m.$$

Some features of (discrete) Martingales

- The *predictable variation process* $\langle M \rangle_n$ for $n > 0$ is the sum of conditional variances of martingale differences,

$$\langle M \rangle_n = \sum_{i=1}^{n} \mathbb{E}\{(M_i - M_{i-1})^2 \mid \mathcal{F}_{i-1}\} = \sum_{i=1}^{n} \text{Var}(\Delta M_i \mid \mathcal{F}_{i-1}),$$

  where $\Delta M_i := M_i - M_{i-1}$. and $\langle M \rangle_0 = 0$.
- The *optional variation process* $[M]_n$ for $n > 0$ is

$$[M]_n = \sum_{i=1}^{n} (M_i - M_{i-1})^2 = \sum_{i=1}^{n} (\Delta M_i)^2,$$

  where $[M]_0 = 0$.

Section 4

Lecture 2

What we did last week

- Introduced hazards and survival functions
- Mentioned censoring
- Looked at the variation processes

Plan for today

- Continue the presentation on Martingales
- Introduce counting processes
- Define censoring assumptions
- Define the multiplicative intensity model
- Study Nelson-Aalen estimator

Take the limits

In continuous time, define:

- the *predictable variation process* $\langle M \rangle_n$ as the limit in probability of the discrete process[4], that is,

$$\langle M \rangle(t) = \lim_{n\to\infty} \sum_{k=1}^{n} \mathrm{Var}(\Delta M_k \mid \mathcal{F}_{(k-1)t/n})$$

  where $[0, t]$ is partitioned into n subintervals of length t/n and $\Delta M_k = M(kt/n) - M((k-1)t/n)$.
  Informally, $\mathcal{F}_{(k-1)t/n} = \mathcal{F}_{t-}$.
- the *optional variation process* $[M]_n$ as

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

ΔM_k are often called "innovations" because, heuristically, they represent what is new and unexpected given the past.

[4] If this limit exists.

Some characteristics

- $M^2 - \langle M \rangle$ is a mean zero martingale.
- $M^2 - [M]$ is a mean zero martingale.
- Thus, $\text{Var}(M(t)) = \mathbb{E}\{M(t)^2\}) = \mathbb{E}\langle M \rangle(t) = \mathbb{E}\{[M](t)\}$.

Submartingale

Definition (Submartingale)

A $\{\mathcal{F}_t\}$-adapted stochastic process $X = \{X(t); t \in [0, \tau]\}$ is a submartingale relative to $\{\mathcal{F}_t\}$ if $\mathbb{E}(X(t) \mid \mathcal{F}_s) \geq X(s)$ for all $t > s$.

That is, $X(t)$ is a process that is expected to increase as time goes on.

Doob-Meyer decomposition

Suppose that X is a submartingale wrt. $\{\mathcal{F}_t\}$.
The Doob-Meyer decomposition theorem states that X can be uniquely decomposed into

$$X = X^* + M,$$

where

- X^* is a non-decreasing predictable process called the "compensator" wrt. $\{\mathcal{F}_t\}$.
- M is a mean zero martingale wrt. $\{\mathcal{F}_t\}$.

We will not show this important result.
However, we will give an argument for discrete processes.

Discrete Doob decomposition

Let $M = \{M_0, M_1, M_2, \dots\}$ be a *discrete* stochastic process adapted to $\{\mathcal{F}_n\}$.
Reminder: the discrete process M is a martingale if

$$\mathbb{E}(M_n \mid \mathcal{F}_{n-1}) = M_{n-1}$$

Now, let $X = \{X_0, X_1, X_2, \dots\}$ be some process with $X_0 = 0$ wrt $\{\mathcal{F}_n\}$, and define $M' = \{M'_0, M'_1, M'_2, \dots\}$ by

$$M'_0 = X_0$$
$$M'_n - M'_{n-1} = X_n - \mathbb{E}(X_n \mid \mathcal{F}_{n-1}).$$

$\Delta M'_n = M'_n - M'_{n-1}$ is a martingale wrt $\{\mathcal{F}_n\}$ because $\mathbb{E}(X_n - \mathbb{E}(X_n \mid \mathcal{F}_{n-1})) = 0)$.
Furthermore,

$$X_n = \mathbb{E}(X_n \mid \mathcal{F}_{n-1}) + \Delta M'_n.$$

Transformation of a martingale

Let $H = \{H_0, H_1, H_2, \dots\}$ 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

$$Z_n = H_0 M_0 + H_1(M_1 - M_0) + \cdots + H_n(M_n - M_{n-1}).$$

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

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

The process Z is denoted the transformation of M by H and it is written $Z = H \bullet M$. There is a close connection to integration here, as we will see.

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

- $N(0) = 0$, $N(t) > 0$, $t \geq 0$,
- $N(t)$ 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

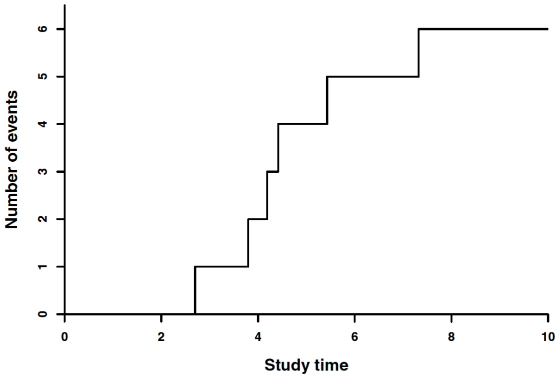


Illustration of a counting process

Using the Doob-Mayer 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. [5] 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.

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

Stochastic integrals for counting processes

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

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

  where $[0, t]$ is partitioned into n subintervals of length t/n and $H_k = H((k-1)t/n)$ and $\Delta M_k = M(kt/n) - M((k-1)t/n)$.[6]
- Importantly, $I(t)$ is a mean zero martingale wrt $\{\mathcal{F}_t\}$.
  Think about this as the limit of discrete time transformations (Slide 40)

[6] While this argument is sufficient for the parameters we study in this course right now, this limiting distribution is not valid in general for stochastic integrals. In more general cases, we would use Itô integrals.

Stochastic integrals for counting processes

From the Doob-Meyer decomposition we also see that

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

- Here, $\int_0^t H(s)dN(s) = \sum_{T_j \leq t} H(T_j)$ is denoted the counting process integral of H, that is, the sum of the values of H at each jump time.

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, 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 $M^2(t)$. 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 (Homework 2)!

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 $T_1, T_2, \ldots, T_n$.[7]
- 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 $N_i^c(t) = I(T_i \leq t)$.[8]
- Define the filtration $\{\mathcal{F}_t^c\}$ is an increasing family of σ algebras generated by $N_i^c(t)$.

[7] 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, $T_i \neq T_j \forall i, j$.

[8] Here superscript "c" denotes complete, to highlight that we have not introduced censoring yet (there is no censoring).

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,

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

- The intensity process $\lambda_i(t)$ 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 λ_i^c for $i = 1, 2, \dots$ on the multiplicative form

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

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

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

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

$$
\begin{aligned}
\lambda^c(t)dt &= \mathbb{E}[dN^c(t)|\mathcal{F}_t^c] \\
&= \sum_{i=1}^n \mathbb{E}[dN_i^c(t)|\mathcal{F}_t^c] \\
&= \sum_{i=1}^n \alpha(t)I(T_i \geq t)dt \text{ when } \alpha_i(t) = \alpha(t).
\end{aligned}
$$

Illustration: Aggregated survival

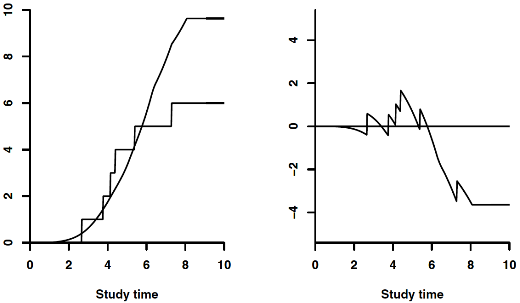


Fig. 1.14 *Illustration of a counting process and its cumulative intensity process (left panel) and the corresponding martingale (right panel).*

Right censoring

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

$$\begin{aligned} \tilde{T}_i &= T_i \;\text{ if }\; D_i = 1, \\ \tilde{T}_i &< T_i \;\text{ if }\; D_i = 0. \end{aligned}$$

- 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 $C_i(t) = I(t \leq T_i^*)$. This process is **left** continuous.
- Let $Z_i(u) = I(t \leq \tilde{T}_i)$.

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)
$\mathcal{F}_0^c = \sigma(L_0)$ and $\mathcal{F}_t^c = \sigma(L_0, N^c(u); 0 \leq u \leq t)$.
$\mathcal{G}_t^c = \sigma(L_0, N^c(u), C(u); 0 \leq u \leq t)$ so $\{\mathcal{G}_t\} \supseteq \{\mathcal{F}_t^c\}$.
$\mathcal{F}_t = \sigma(L_0, N(u), Z(u); 0 \leq u \leq t)$.

Independent censoring

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.

Independent censoring allows us to write....

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

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

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

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

where $Z_i^c(t) = I(t \leq T_i)$.
Thus, importantly, we can identify α_i from the **censored** data under independent censoring.