2

An informal introduction to hazard-based analyses

This chapter explains in a non-technical manner why methods for analysing standard survival data — one endpoint, observation of which is subject to right-censoring — transfer to more complex models, namely competing risks and multistate models, this book's topic.

In Section 2.1, we explain how right-censoring, where only an individual's minimum lifetime may be known due to closing of a study, say, leads to the hazard rate being the key statistical quantity. Probability estimates are then derived as deterministic functions of simpler estimators of the cumulative hazards. In fact, the technical difficulty of right-censoring is a consequence of an important conceptual aspect: time is not just another measurement on a scale, but plays a special role. Events happen over the course of time, e.g., illness often precedes death, and one has to wait in order to observe an event. This requires a dedicated statistical theory, and hazards are in general well suited to analyse events that occur over the course of time like survival data do. The concept of a hazard is also important for the data analyst's intuition: approximately, one may think of a hazard as the probability of experiencing an event within the next time unit conditional on presently being event-free. Say the event is death. Then this information may be more relevant given current vital status than an unconditional survival probability.

In Section 2.2, we explain why hazard-based techniques also apply to analysing competing risks data and multistate model data. Competing risks models allow for investigating different endpoint types that may occur at the event time in question. Occurrence of subsequent events may be investigated by multistate models. The type of multistate models that we consider are time-inhomogeneous Markov models, which are realized as a series of nested competing risks experiments. It is also discussed that these techniques allow for left-truncation in addition to right-censoring. Data are left-truncated if patients have a delayed study entry time. The mathematical basis behind these extensions are counting processes, which count different event types over the course of time, and martingales, which represent noise over the course of time.

J. Beyersmann et al., *Competing Risks and Multistate Models with R*, Use R,
DOI 10.1007/978-1-4614-2035-4_2, © Springer Science+Business Media, LLC 2012

The connection to counting processes and martingales is also explained in an informal way.

Finally, the brief Section 2.3 explains that asymptotic results are used for approximate inference when analysing event times in practice. Typically, the statistical techniques are of a nonparametric kind, and, e.g., using approximate normality has been found to work well even in moderate sample sizes.

Section 2.1 contains a lot of R code. Readers are encouraged to reproduce the code which explains how we may estimate cumulative hazards in the presence of right-censoring and how probabilities and their estimates may be computed from either the true or the estimated cumulative hazards. There is hardly any R code in Sections 2.2 and 2.3; this is what the remainder of the book is about.

2.1 Why survival analysis is hazard-based.

2.1.1 Survival multistate model, hazard, and survival probability

Figure 2.1 displays the simplest multistate model. An individual is in the initial

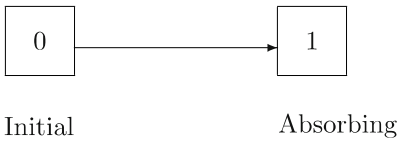

Fig. 2.1. Survival multistate model.

state 0 at time origin. At some later random time T, the individual moves to the absorbing state 1. 'Absorbing' means that the individual cannot move out of state 1, or that transitions out of state 1 are not modelled. Figure 2.1 is the classical model of survival analysis, if we interpret state 0 as 'alive' and state 1 as 'dead'. We are interested in the event time T; T is often called 'survival time' or 'failure time'.

To find out about T, we need to record data over the course of time, i.e. we need to record in which state, 0 or 1, an individual is for every point in time. This is what a stochastic process does. We write X_t for the state occupied by the individual at time $t \geq 0$, $X_t \in \{0, 1\}$. T is the smallest time at which the process is not in the initial state 0 anymore,

$$T := \inf\{t \,:\, X_t \neq 0\}. \tag{2.1}$$

This relationship between the stochastic process $(X_t)_{t\geq 0}$ and the event time T is illustrated in Figure 2.2. For illustration, consider an individual with event time $T = 52$. This individual will be in state 0 for all times $t \in [0, 52)$ and in

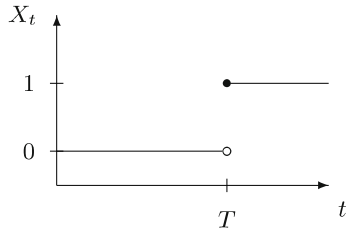

Fig. 2.2. Stochastic process $(X_t)_{t\geq 0}$ and the event time T: The bullet • is included in the graph, and the circle ∘ is not.

state 1 for all times $t \geq 52$. Note that the state occupied at the event time T is the absorbing state 1, i.e., $X_T = X_{52} = 1$. This definition implies that the sample paths of the stochastic process, i.e.

$$[0, \infty) \ni t \mapsto X_t$$

are right-continuous, as illustrated in Figure 2.2. As state 1 is absorbing, data recording may stop for an individual with time T.

The statistical analysis of T is based on the hazard $\alpha(t)$ attached to the distribution of T:

$$\alpha(t) \cdot \mathrm{d}t := \mathrm{P}(T \in \mathrm{d}t \,|\, T \geq t), \tag{2.2}$$

where we write $\mathrm{d}t$ both for the length of the infinitesimal (i.e., very small) time interval $[t, t + \mathrm{d}t)$ and the interval itself. Equation (2.2) is a short, but more intuitive form of

$$\alpha(t) := \lim_{\Delta t \searrow 0} \frac{\mathrm{P}(T \in [t, t + \Delta t) \,|\, T \geq t)}{\Delta t}.$$

Throughout this book, we assume that derivatives such as in Equation (2.2) exist. The hazard is 'just' a different 'representation' of the distribution of T:

Before we answer the question of why survival analysis is hazard-based, let us first note that knowing the cumulative hazard $A(t)$,

$$A(t) := \int_0^t \alpha(u)\,\mathrm{d}u, \tag{2.3}$$

suffices to recover the distribution function of T,

$$F(t) := 1 - S(t) := \mathrm{P}(T \leq t) = 1 - \exp(-A(t)), \tag{2.4}$$

where $S(t) = \mathrm{P}(T > t) = \exp(-A(t))$ is usually called the survival function of T. The right hand side of (2.4) is easily derived from (2.2) using standard calculus, but this adds little to our understanding. A more useful notion is product integration: because $\mathrm{d}A(u) = \alpha(u)\mathrm{d}u$, we may rewrite (2.2),

$$1 - \mathrm{d}A(u) = \mathrm{P}(T \geq u + \mathrm{d}u \,|\, T \geq u). \tag{2.5}$$

The survival function should then be an infinite product over conditional probabilities of type (2.5). This is, in fact, the case. We call such an infinite product a product integral and write $\prod$. So,

$$S(t) = \prod_0^t (1 - \mathrm{d}A(u)) \tag{2.6}$$

$$\approx \prod_{k=1}^{K} (1 - \Delta A(t_k)) \approx \prod_{k=1}^{K} \mathrm{P}(T > t_k \,|\, T > t_{k-1}), \tag{2.7}$$

where $0 = t_0 < t_1 < t_2 < \ldots < t_{K-1} < t_K = t$ partitions the time interval $[0, t]$ and $\Delta A(t_k) = A(t_k) - A(t_{k-1})$. Now, the right hand side of (2.4) can simply be seen as a solution of the product integral in (2.6). The product integral itself, however, shows up again with the famous Kaplan-Meier or product limit estimator of the survival function, and, in a matrix-valued form, when we move from survival analysis to competing risks and multistate models. Let us now check the approximation of (2.6) by (2.7) empirically using R. The following function `prodint` takes a vector of time points and a cumulative hazard A as an argument and returns the approximation $\prod_{k=1}^{K} (1 - \Delta A(t_k))$.

```
> prodint <- function(time.points, A) {
+     times <- c(0, sort(unique(time.points)))
+     S <- prod(1 - diff(apply(X = matrix(times),
+                         MARGIN = 1, FUN = A)))
+     return(S)
+ }
```

A standard parametric example is the exponential distribution with constant hazard $\alpha(t) = \alpha$ and cumulative hazard $A(t) = \alpha \cdot t$. We exemplarily look at an exponential distribution with hazard 0.9,

```
> A.exp <- function(time.point) {
+     return(0.9 * time.point)
+ }
```

on the time interval $[0, 1]$:

```
> times <- seq(0, 1, 0.001)
> prodint(times, A.exp)

[1] 0.4064049

> exp(-0.9 * max(times))

[1] 0.4065697
```

The vector of time points does not have to be equally spaced:

```
> prodint(runif(n = 1000, min = 0, max = 1), A.exp)
```

```
[1] 0.4068678
```

A more flexible parametric model is the Weibull distribution with shape parameter θ and scale parameter γ. It has hazard $\alpha(t) = \gamma \cdot \theta \cdot t^{\theta-1}$ and cumulative hazard $A(t) = \gamma \cdot t^{\theta}$. Let us look at a Weibull distribution with scale 2 and shape 0.25,

```
> A.weibull <- function(time.point){
+   return(2 * time.point^0.25)
+ }
```

and the time interval $[0, 1]$ as before:

```
> prodint(times, A.weibull)

[1] 0.1234838

> exp(-2 * max(times)^0.25)

[1] 0.1353353
```

The approximation becomes better with an ever finer spaced partition:

```
> prodint(seq(0, 1, 0.000001), A.weibull)

[1] 0.1350193
```

The next section 2.1.2 explains why the cumulative hazard can still be estimated, if data are incomplete due to, e.g., individuals surviving the closing of a study. An estimator of the survival probability is then derived by computing the product integral with respect to the estimated cumulative hazard.

2.1.2 Estimation: The hazard remains 'undisturbed' by censoring.

The approximation (2.7) of the product integral (2.6), implemented via `prodint`, directly results in the Kaplan-Meier estimator of $S(t)$, if we substitute the increment of the cumulative hazard by an adequate estimator, which turns out to be the Nelson-Aalen estimator of the cumulative hazard. In other words, we may estimate the survival function S of the event time T by the product integral of an estimator of the cumulative hazard. This leads us to the question of why survival analysis is hazard-based, and, of course, how the cumulative hazard may be estimated: so far, we may estimate $S(t)$ either by the empirical survival function,

$$n^{-1} \cdot (\text{number of individuals surviving } t)\,, \tag{2.8}$$

if we start with n individuals at time origin. Or we may base things on estimating the cumulative hazard. A natural estimator of the increments $\Delta A(t)$ is

$$\Delta\widehat{A}(t) = \frac{\text{number of individuals failing at } t}{\text{number of individuals alive just prior to } t}, \tag{2.9}$$

so that we estimate the cumulative hazard as

$$\widehat{A}(t) = \sum_{k=1}^{K} \frac{\text{number of individuals failing at } t_k}{\text{number of individuals alive just prior to } t_k}, \tag{2.10}$$

if $0 < t_1 < t_2 < \ldots < t_{K-1} < t_K = t$ is the ordered sequence of observed failure times. It is a straightforward algebraic exercise to show that $\prod_0^t \left(1 - \mathrm{d}\widehat{A}(u)\right) = \prod_{k=1}^K \left(1 - \Delta\widehat{A}(t_k)\right)$ equals (2.8), if we observe the failure times for all n individuals. We briefly check this with R: we simulate 100 independent random variables from the exponential distribution with parameter 0.9,

```
> event.times <- rexp(100,0.9)
```

for which we wish to estimate the survival distribution at $t = 1$. We now need to compute the increments (2.9), which can be conveniently done using the `survival` package (Therneau and Grambsch, 2000): First, we create the fundamental 'survival object' using `Surv` on the simulated times. The `survfit`-function then gives us the necessary information:

```
> library(survival)
> fit.surv <- survfit(Surv(event.times) ~ 1)
```

Now, `fit.surv$time` is the vector of times $t_1 < t_2 < \ldots < t_{K-1} < t_K$, and `fit.surv$n.event` and `fit.surv$n.risk` are the numerator and denominator of (2.9), respectively. The function `A` computes (2.10):

```
> A <- function(time.point) {
+     sum(fit.surv$n.event[fit.surv$time <= time.point]/
+         fit.surv$n.risk[fit.surv$time <= time.point])
+ }
```

The estimator of the survival function at time 1 based on `A` and product integration then is

```
> prodint(event.times[event.times <= 1], A)
```

```
[1] 0.41
```

and the empirical survival function is

```
> sum(event.times > 1) / length(event.times)
```

```
[1] 0.41
```

The restrictive assumption required to use the empirical survival function (2.8) is that we are supposed to know the actual failure times of all individuals. This will usually not be the case. Event history data occur over the course of time,

and a data analysis is regularly performed before or without knowing all failure times. E.g., a clinical study may be closed with, one hopes, many patients surviving, or individuals may drop out of a study because they move to a different place. In these instances, we will only know the minimum failure time. That is, we only know the actual failure time to be greater than a certain value, but not its precise value. This mechanism leads to incomplete observations and is known as (right-)censoring. In the presence of censoring, the empirical survival function (2.8) is rendered useless, as we cannot compute it anymore. *But hazards remain undisturbed by censoring:* recall Definition (2.2) of the hazard. Now introduce a censoring time C, independent of the event time T. (This is the so-called random censorship model.) The observation is

$$(T \wedge C, \mathbf{1}(T \leq C)), \tag{2.11}$$

where we write $\wedge$ for the minimum and $\mathbf{1}(\cdot)$ for the indicator function: $\mathbf{1}(T \leq C)$ equals 1, if T is less than or equal to C. $T \wedge C$ is the censored event time, and the event indicator $\mathbf{1}(T \leq C)$ tells us, whether $T \wedge C$ equals the actual event time T. Now, what is the probability of observing the actual event time in the small time interval $\mathrm{d}t = [t, t+\mathrm{d}t)$, conditional on the fact that neither event nor censoring have happened before t? I.e., what is $\mathrm{P}(T \in \mathrm{d}t, T \leq C \,|\, T \wedge C \geq t)$? The interval $\mathrm{d}t$ is so short that, assuming T and C to be different, at most one is in $\mathrm{d}t$: if the event occurs in $\mathrm{d}t$, it will be observed (still supposing $T \wedge C \geq t$). Because C and T are independent, the probability that the event occurs in $\mathrm{d}t$, conditional on $T \wedge C \geq t$, is the same as in the absence of censoring:

$$\alpha(t) \cdot \mathrm{d}t = \mathrm{P}(T \in \mathrm{d}t \,|\, T \geq t) = \mathrm{P}(T \in \mathrm{d}t, T \leq C \,|\, T \wedge C \geq t). \tag{2.12}$$

In words: censoring has not disturbed the hazard. As a consequence, we may estimate the cumulative hazard from censored data. Using product integration, this results in an estimator of the survival function. Equation (2.12) has farther reaching consequences, which we investigate in Section 2.2. These are also seen to be the reason why hazard-based techniques translate from the simple survival multistate model of Figure 2.1 to competing risks and more complex multistate models. Before we do so, let us briefly investigate estimation from censored data:

We say an individual with $T \wedge C \geq t$ is 'at risk' just prior to t. In order to estimate the cumulative hazard, adapting Equations (2.9) and (2.10) to the censored data set-up is straightforward:

$$\Delta\widehat{A}(t) = \frac{\text{number of individuals } \textit{observed} \text{ to fail at } t}{\text{number of individuals } \textit{at risk} \text{ just prior to } t}, \tag{2.13}$$

so that we estimate the cumulative hazard as

$$\widehat{A}(t) = \sum_{k=1}^{K} \frac{\text{number of individuals } \textit{observed} \text{ to fail at } t_k}{\text{number of individuals } \textit{at risk} \text{ just prior to } t_k}, \tag{2.14}$$

if $0 < t_1 < t_2 < \ldots < t_{K-1} < t_K = t$ is the ordered sequence of observed failure times. $\widehat{A}$ of (2.14) is the Nelson-Aalen estimator of the cumulative hazard. The product integral of $\widehat{A}$ is the Kaplan-Meier estimator of the survival function:

$$\widehat{S}(t) := \prod_0^t \left(1 - \mathrm{d}\widehat{A}(u)\right) = \prod_{k=1}^{K} \left(1 - \Delta\widehat{A}(t_k)\right) \tag{2.15}$$

We briefly revisit the R data example from above: in addition to `event.times`, we simulate censoring times, which we choose to be uniformly distributed on $[0, 5]$:

```
> cens.times <- runif(100,0,5)
```

The observable data are the censored event times $T \wedge C$,

```
> obs.times <- pmin(event.times, cens.times)
```

and the event indicator $\mathbf{1}(T \leq C)$,

```
> event.times <= cens.times
```

The number of observed event times is

```
> sum(event.times <= cens.times)

[1] 80
```

We now have to refit the survival object, also telling `Surv` which event times were observed and which were censored:

```
> fit.surv <- survfit(Surv(obs.times,
+                          event.times <= cens.times) ~ 1)
```

The Kaplan-Meier estimator of the survival function at time 1 then is

```
> prodint(obs.times[obs.times<=1], A)

[1] 0.3967501
```

The result is reasonably close both to the estimate previously obtained in the absence of right-censoring and to the true value. Of course, we may also use the `survival` package in order to estimate the survival function at time 1:

```
> S <- fit.surv$surv
> S[fit.surv$time <= 1][length(S[fit.surv$time <= 1])]

[1] 0.3967501
```

So far, we have only estimated the survival function at one time point. Evaluating formula (2.15) at all observed event times yields an estimate of the survival *curve*. A plot of the estimated survival function together with its theoretical counterpart is displayed in Figure 4.3 in Section 4, where a more in-depth discussion of the Kaplan-Meier estimator, the `survival` package, and plotting the respective results is given.

In summary, survival analysis is hazard-based, because we can still estimate the cumulative hazard from right-censored data. We may then use product integration to recover the survival function or, equivalently, the distribution function. In the remainder of this chapter, we find that this program still works, in essence, with even more complex event data. Crucial to this is an intimate relationship between hazards and counting processes; the latter do a very intuitive thing: they count the number of observed events of a certain type over the course of time. However, the *interpretation* of hazard-based results becomes more involved with more complex event data, which is a major topic of this book.

2.2 Consequences of survival analysis being based on hazards

In Section 2.1, we illustrated that the analysis of event time data is based on hazards. This fact has a number of important consequences, which are briefly outlined below. In Section 2.2.1, we find that estimation of the cumulative hazard is intimately connected to counting processes and martingales. A counting process simply counts the number of observed events over the course of time. Martingale theory provides us with estimating equations and both small and large sample properties of estimators. This connection allows us to also analyse event time data which go beyond the right-censored, single-event type situation discussed in Section 2.1.

In Section 2.2.2, we show that the hazard-based approach can also account for left-truncated data, where patients have delayed study entry times. Sections 2.2.3 and 2.2.4 show how the current framework generalizes to competing risks and to time-inhomogeneous Markov multistate models. In addition to considering an event time, competing risks models also distinguish between different event types, one of which occurs at the event time. Multistate models can be thought of as being realized as a series of nested competing risks experiments: an individual may experience different events over the course of time, which are modelled as transitions between multiple states.

2.2.1 Counting processes and martingales

In Equation (2.12), we found that random right-censoring does not disturb the hazard. We reformulate (2.12) as

$$\mathrm{E}\left(\mathbf{1}(T \in \mathrm{d}t, T \leq C) \,|\, \text{Past}\right) = \mathbf{1}(T \wedge C \geq t) \cdot \alpha(t)\, \mathrm{d}t, \tag{2.16}$$

where 'Past' stands for knowledge about all failure or censoring events before t. Given the past, the at-risk indicator $\mathbf{1}(T \wedge C \geq t)$ is known. If the individual is at risk just prior to t (i.e., if we have that $T \wedge C \geq t$), the probability (conditional on the past) that we observe an event in the very small time interval $\mathrm{d}t$ is as in (2.12). However, if the individual is not at risk just prior to t because either a failure or a censoring event has happened before t (i.e., $\mathbf{1}(T \wedge C \geq t) = 0$), this probability is zero. This is summarized in (2.16).

As outlined earlier, Equation (2.12) implies that the cumulative hazard is estimable from the observable data. In fact, (2.12) shows that a key role in the estimation is played by the counting process

$$t \mapsto \mathbf{1}(T \leq t, T \leq C), t \geq 0, \tag{2.17}$$

which has (infinitesimal) increments $\mathbf{1}(T \in \mathrm{d}t, T \leq C)$. The process (2.17) simply counts the number of observed events in the time interval $[0, t]$, either 0 or 1. Attached to the counting process is the at-risk process

$$t \mapsto \mathbf{1}(T \wedge C \geq t). \tag{2.18}$$

The counting process may only jump (from 0 to 1) at time t, if the at-risk process equals 1. For estimation, we aggregate these processes over all individuals under study such that we count the number of observed events within the sample and over the course of time. The at-risk process then keeps track of the number of individuals currently at risk, i.e., without prior failure or censoring event. This is reflected in the Nelson-Aalen estimator (2.14).

Furthermore, Equation (2.12) is tantamount to the fact that

$$\mathbf{1}(T \leq t,\, T \leq C) - \int_0^t \mathbf{1}(T \wedge C \geq u) \cdot \alpha(u)\, \mathrm{d}u \tag{2.19}$$

is a so-called martingale. Martingale theory provides a powerful tool to derive estimators and test statistics as well as to study their small and large sample properties. The latter may be used for approximate inference in practice. An in-depth treatment of the application of martingale theory to the analysis of event time data is beyond the technical level of this book. Interested readers are referred to Andersen et al. (1993) and Aalen et al. (2008). We, however, often make use of the results provided by the application of martingale theory. E.g., variance estimators and approximate 95% confidence intervals may conveniently be derived in this way.

We note, though, that equations (2.16) and (2.19), aggregated over all individuals under study, suggest estimating $A(t) = \int_0^t \alpha(u)\, \mathrm{d}u$ by the Nelson-Aalen estimator

$$\widehat{A}(t) = \sum_{k=1}^{K} \frac{\text{number of individuals observed to fail at } t_k}{\text{number of individuals at risk just prior to } t_k},$$

where the summation is over all event times t_k, which are less than or equal to t (cf. (2.14)). Equation (2.16) also suggests that $\widehat{A}$ should be an almost unbiased estimator of A, as long as the risk set (i.e., the set of all individuals currently at risk) is non-empty with a high probability. Martingale theory can be used to show that this is actually the case.

In other words, the martingale (2.19), potentially aggregated over all individuals, can be considered as a noise process. Figure 2.3 shows the counting process

$$t \mapsto \text{number of individuals observed to fail in } [0, t]$$

computed based on the simulated data in Section 2.1 and its so-called compensator

$$t \mapsto \int_0^t (\text{number of individuals at risk just prior to } u) \cdot 0.9 \, \mathrm{d}u,$$

i.e., the integral over the at-risk process times the hazard $\alpha(t) = 0.9$ of the uncensored event times (cf. equations (2.16) and (2.19)). Figure 2.3 illustrates

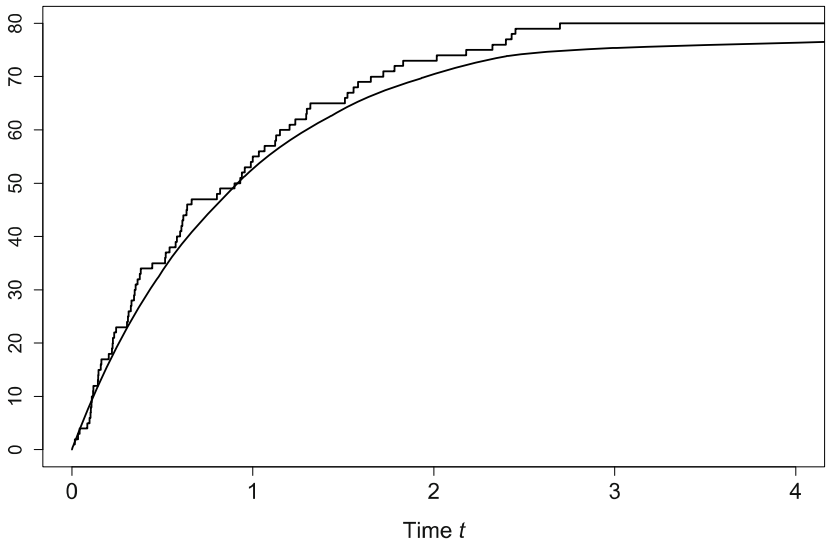

Fig. 2.3. *Simulated data.* The step function is the counting process of observed events. The smooth line is the compensator of the counting process, i.e. the integral over the risk set times the true hazard. Note that the compensator is only an almost smooth line which has 'edges'. The left hand derivative does not equal the right hand derivative at time points where the number of individuals at risk changes.

that the counting process approximates the compensator, which is intimately related to the cumulative hazard. The martingale itself, i.e., the counting process minus the compensator, is unpredictable zero-mean noise.

Readers are encouraged to check the approximation illustrated in Figure 2.3 in R: as explained in Section 2.1, the increments of the counting process at times `fit.surv$time` are contained in `fit.surv$n.event`, which allows for convenient computation of the counting process itself. The integral over the at-risk process times 0.9 may be computed from `fit.surv$n.risk`, which contains the number of individuals at risk just prior to times `fit.surv$time`.

The fact that counting the number of observed events over the course of time approximates a quantity which is closely related to the cumulative hazard, which, in turn, is a key target quantity for estimation, makes counting processes and martingales a starting point for a rich statistical theory. Sections 2.2.2–2.2.4 discuss how we can profit from the counting process approach for more complex event time patterns.

2.2.2 Left-truncation and right-censoring

So far, we have considered the situation where observation of an event time T is restricted by a right-censoring time C: if the event happens in $(0, C]$, it will be observed. However, if the event happens after C, we will only know that T exceeds C. A typical example is a clinical study where the time origin 0 corresponds to random assignment of a patient to a treatment. The event time T then measures the patient's survival time since treatment assignment. Patients who survive beyond administrative closing of the study will be right-censored. Usually, the study is closed at a particular fixed date such that we may assume the right-censoring time C to be independent of the event time T, random right-censorship. In this set-up, individuals are assumed to be followed from time 0 until $T \wedge C$.

There are, however, situations where individuals enter the study at a time later than time origin 0. Such data with a delayed study entry time are said to be left-truncated. The concept is best understood via an example. Meister and Schaefer (2008) study duration of drug-exposed pregnancies. Observation does not start at time of conception. In the study of Meister and Schaefer, women enter the study when first contacting a Teratology Information Service. For these women, the time of conception may reasonably well be determined in retrospect and is thus assumed to be known. However, women who, e.g., experience a spontaneous abortion before their potential future study entry time never enter the study. The hazard/counting process based approach of Section 2.2.1 also allows us to analyse such left-truncated data. The data analysed in Meister and Schaefer (2008) are available in the R package `etm`.

In addition to T and C, we denote an individual's left-truncation/study entry time by L: the event will only be observed, if it happens in $(L, C]$. If it happens in $(L, C]$, we will know T. An individual who experiences an

event before its left-truncation time (i.e., $T \leq L$) will never enter the study. An individual under study (i.e., an individual with $L < T$) is right-censored, if it experiences an event after its right-censoring time (i.e., $C < T$). Data subject to right-censoring only are included in this set-up: they formally have a left-truncation time $L = 0$. Similarly, data which are only subject to left-truncation formally have a right-censoring time beyond the largest possible event time.

Potential subsequent occurrences of L, T, and C are schematically illustrated in Figure 2.4. The individual in Figure 2.4 a) enters the study at time L

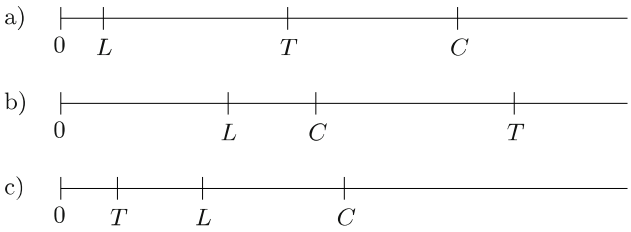

Fig. 2.4. a) Individual with an observed event. b) Censored individual. c) Individual with an event before study entry.

after time origin 0. The individual's event is observed at time T, because the censoring time C is larger than T. The individual in Figure 2.4 b) also enters the study at a time L, but the observation of the individual is right-censored. In Figure 2.4 c), the individual experiences an event before its study entry time, i.e., $T < L$. This individual never enters the study.

We assume for the moment random right-censoring and random left-truncation: T is independent of (L, C). Then Equation (2.16) generalizes to

$$\mathrm{E}\left(\mathbf{1}(T \in \mathrm{d}t, L < T \leq C) \,|\, \text{Past}\right) = \mathbf{1}(L < t \leq T \wedge C) \cdot \alpha(t)\,\mathrm{d}t, \tag{2.20}$$

where 'Past' now means knowledge about all failure, truncation, or censoring events before t. As in (2.16), Equation (2.20) states that the probability of observing an event in $\mathrm{d}t$ is 0, if the individual is not at risk just before t. However, if the individual is at risk, an event that happens in $\mathrm{d}t$ will be observed; such an event occurs with probability $\alpha(t)\,\mathrm{d}t$. The key point is that the at-risk process

$$t \mapsto \mathbf{1}(L < t \leq T \wedge C) \tag{2.21}$$

now also accounts for left-truncation. An individual is only 'at risk' right after the time L of study entry. The Nelson-Aalen estimator (2.14)

$$\widehat{A}(t) = \sum_{k=1}^{K} \frac{\text{number of individuals observed to fail at } t_k}{\text{number of individuals at risk just prior to } t_k},$$

of the cumulative hazard $A(t) = \int_0^t \alpha(u)\,\mathrm{d}u$ is straightforwardly adapted to data subject to both left-truncation and right-censoring. The denominator now includes all individuals who have entered the study before t_k but have not experienced an actual event or a censoring event before t_k. The numerator counts the number of observed events at t_k within the set of these individuals.

A further important consequence of Equation (2.20) is that it allows us to relax the assumption of random right-censoring and random left-truncation. The application of martingale theory and counting processes only requires Equation (2.20) to hold, but not necessarily random right-censoring or random left-truncation. Restrictions of observing T that fulfill (2.20) are known as independent right-censoring and independent left-truncation, respectively. A crucial issue here is that L and C may depend on the 'Past': if covariates are considered in the statistical analysis, L and C may depend on past covariate values. A simple example of this is a clinical study where censoring may differ between treatment groups; see, e.g., Clark et al. (2002). A more complex example is the illness-death model as considered in Section 2.2.4 below. In this model, individuals may experience different events over time, i.e., undergo 'healthy' $\leftrightarrow$ 'diseased' transitions and may also die. A right-censoring mechanism that is independent censoring in the aforementioned sense may depend on whether an individual is currently 'healthy' or 'diseased'. Because of this potential dependency, however, such right-censoring would not be random any more.

So far, our discussion of left-truncation and right-censoring referred to situations where observation of an individual was restricted due to some 'external' mechanism. Observation does not start before the time of left-truncation, and an event that happens after the right-censoring time will not be observed. In Sections 2.2.3 and 2.2.4, our aim is to estimate multiple (cumulative) hazards that correspond to multiple event types. Equation (2.20) can be generalized to such situations. A key point is to adapt the risk sets. In the aforementioned illness-death model, individuals will only be at risk of making a 'diseased' $\rightarrow$ 'dead' transition, after having acquired the disease and thus having entered the disease state. The set of individuals at risk of making a 'healthy' $\rightarrow$ 'diseased' transition not only excludes individuals who have previously fallen ill and have not yet recovered, but also those who previously died without prior disease. As discussed below, the appropriate changes to the risk set may be made by coding these as left-truncation and right-censoring, respectively, but these modifications of the risk set are due to the presence of multiple event types and not to external restrictions on an individual's observable data. We note, however, that, conversely, the presence of multiple event types may motivate data collection which is subject to left-truncation. E.g., in hospital epidemiology, the time scale of interest typically is time since admission to hospital,

but sometimes data are collected conditional on detection of some infectious strain during hospital stay (Beyersmann et al., 2011).

In brief, left-truncation corresponds to observation being switched on, and right-censoring corresponds to observation being switched off. If these switches, which are, in fact, a 'censoring' process, are independent as explained above, 'the Nelson-Aalen estimator works'. We make two final comments on this: First, the idea of switching observation on and off may be generalized to quite complex observation schemes called filtering, and it disposes of the latent variables L and C. We do not further pursue the concept of filtering here. The variables L and C are latent in the sense that, e.g., L is unobservable if $L \geq T$, which is somewhat unpleasant. In contrast, the concept of 'observation on' as mirrored in the risk set does not require these latent times. In fact, our discussion of the illness-death model above has implicitly used the concept of 'observation on/off' rather than latent times. We have, however, chosen to use L and C in line with many accounts in the applied literature. Second, the independence assumption is obviously crucial, and it would therefore be useful if one could check it for a real data set. Unfortunately, there are identifiability problems that typically prevent checking the assumption for right-censoring, but it may be investigated for left-truncation. For the latter, see Section 11.3.

We finally note that left-truncation should not be confused with left-censoring. For a left-censored event time, we know that an event has happened before a left-censoring time in the past, but the exact time is not known. E.g., the occurrence time of a certain disease is left-censored, if it is only known to have occurred before the time of diagnosis. Klein and Moeschberger (2003) give a very readable account of the different variants of truncation and censoring. Our current approach conveniently allows for right-censoring, which is the most frequent reason for incompletely observed event time data, and left-truncation.

2.2.3 Competing risks

So far, we have considered a time T until one single possible event. The standard example is time until death, hence the name survival analysis. Often, however, a combined endpoint is considered. E.g., medical studies often investigate 'disease-free survival', i.e., time until (recurrence of a) disease or death (without prior disease), whatever comes first. In economics, one might wish to study durations of unemployment, ended either by finding a new job or retirement. Thus, T in general denotes time until some first event. The aim of a competing risks model is to distinguish between the possible types of that first event.

The analysis of competing risks is covered in depth in Chapters 3–7. At the current stage, the key question is how to generalize the basic two-state survival model of Figure 2.1 to competing risks. Our preceding discussion implies that an individual moves into the absorbing state of Figure 2.1 at time T, when the first of the possible events under study occurs. In other words, the

absorbing state of Figure 2.1 represents a combined endpoint. The two-state survival model may now be generalized to competing risks by introducing several competing absorbing states which represent the possible event types. Occurrence of a competing event is modelled by a transition into the corresponding competing event state. Such a model is depicted in Figure 2.5 and a finite number J of competing risks. Figure 3.1 in Chapter 3 displays the corresponding model for two competing risks.

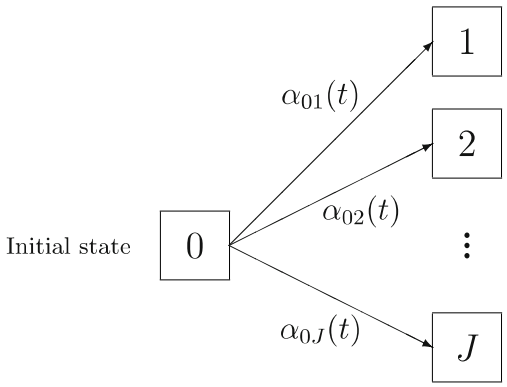

Fig. 2.5. Competing risks multistate model with cause-specific hazards $\alpha_{0j}(t)$, $j = 1, 2, \ldots, J$. The vertical dots indicate the competing event states $3, 4, \ldots, J-1$.

Restricting for the moment the discussion to two competing risks, the stochastic process $(X_t)_{t\geq 0}$ attached to Figure 2.1 may easily be extended to the competing risks setting. Again, X_t simply denotes the state occupied by the individual at time $t \geq 0$. X_t equals 0, if the individual is still event-free at time t. Coding, as in Figure 3.1, the two potential competing events as 1 and 2, X_t equals 1, if event type 1 has occurred in $[0, t]$. If event type 2 has occurred in $[0, t]$, $X_t = 2$. As before, the event time T is the smallest time at which the process is not in the initial state 0 anymore; $T := \inf\{t : X_t \neq 0\}$ (cf. Equation (2.1)).

In addition to the event time T, competing risks data consist of a second component, the event type. Recall from our discussion of Figure 2.2 that X_T denotes the absorbing state entered at time T. In our setting with two competing event states, X_T equals either 1 or 2. I.e., X_T denotes the event type, and complete competing risks data consist of the tuple (T, X_T). More than two competing risks are easily included in this set-up by letting $X_T \in \{1, 2, \ldots, J\}$.

As illustrated in Figure 3.1, we now have one event-specific hazard per competing event. Paralleling Definition (2.2), these are defined as

$$\alpha_{0j}(t) \cdot \mathrm{d}t := \mathrm{P}(T \in \mathrm{d}t, X_T = j \,|\, T \geq t),\ j = 1, \ldots, J, \tag{2.22}$$

where the index $0j$ denotes the transition type out of the initial state 0 into the competing event state j. The α_{0j}s are often called cause-specific hazards (e.g., Prentice et al., 1978). The interpretation of (2.22) is that $\alpha_{0j}(t) \cdot \mathrm{d}t$ is the probability that a type j event happens in the small time interval $\mathrm{d}t = [t, t + \mathrm{d}t)$, conditional on the fact that no event (of any type) has happened before t. For the more complex multistate models considered in Section 2.2.4 below, it is useful to rewrite Definition (2.22) in terms of the simple stochastic process $(X_t)_{t \geq 0}$,

$$\alpha_{0j}(t) \cdot \mathrm{d}t = \mathrm{P}(X_{(t+\mathrm{d}t)-} = j \,|\, X_{t-} = 0),\ j = 1, \ldots, J, \tag{2.23}$$

where X_{t-} denotes the state occupied just before time t.

As in Section 2.2.2, we call mechanisms of left-truncation and right-censoring independent, if they do not change these probabilities, i.e.,

$$\mathrm{E}\left(\mathbf{1}(T \in \mathrm{d}t, X_T = j, L < T \leq C) \,|\, \mathrm{Past}\right) = \mathbf{1}(L < t \leq T \wedge C) \cdot \alpha_{0j}(t)\, \mathrm{d}t, \tag{2.24}$$

$j = 1, \ldots, J$. A cause-specific Nelson-Aalen estimator of the cumulative hazard $A_{0j}(t) = \int_0^t \alpha_{0j}(u)\, \mathrm{d}u$ is now given as

$$\widehat{A}_{0j}(t) = \sum_{k=1}^{K} \frac{\text{number of observed type } j \text{ events at } t_k}{\text{number of individuals at risk just prior to } t_k}, \tag{2.25}$$

$j = 1, \ldots, J$, where the summation is over all event times t_k, which are less than or equal to t. In Chapter 3, we show that we can compute probability estimates as deterministic functions of $t \mapsto (\widehat{A}_{01}(t), \ldots, \widehat{A}_{0J}(t))$. Here, we note two important facts which have already been alluded to earlier: first, the numerator in (2.25) now represents increments of a cause-specific counting process. Second, the risk set in (2.25) excludes all prior type j events, all prior censoring events, and all prior events of a type $\tilde{\jmath}$, $\tilde{\jmath} \neq j$.

In other words, when coding computation of $\widehat{A}_{0j}(t)$, say, we may code type $\tilde{\jmath}$ events, $\tilde{\jmath} \neq j$, as censoring events and only count type j events as 'actual events': occurrence of type $\tilde{\jmath}$ events acts as independent right-censoring with respect to type j events. This means that removal of prior type $\tilde{\jmath}$ events from the risk set allows for estimation of $A_{0j}(t) = \int_0^t \alpha_{0j}(u)\, \mathrm{d}u$. However, in Chapter 3, we also show that such 'censoring by a competing event' is informative in the sense that probability estimates depend on computing all $\widehat{A}_{01}(t), \ldots, \widehat{A}_{0J}(t)$. This has two important implications for any competing risks analysis:

- In a cause-specific hazards analysis, competing events may be coded as a censoring event.
- This has to be done for every competing event type in turn.

These two steps are, in particular, illustrated in Chapter 5.

Finally, in Section 2.2.4 below, we extend the hazard-based approach to multistate models, which can be thought of as being realized as successive nested competing risks experiments. To this end we note that competing risks data can be considered as realizations of a two-step simulation experiment that determines the time T at which the event occurs via the all-cause hazard $\alpha(t) = \alpha_{01}(t) + \ldots \alpha_{0J}(t)$ (i.e., the usual hazard of the event time T); the event type X_T for a given time T is determined via a multinomial experiment that decides with probability $\alpha_{0j}(T)/\alpha(T)$ on $X_T = j$. This simulation point of view towards competing risks data shows up again and again in the main part of the book. See, in particular, Sections 3.2 and 5.2.2 for an in-depth treatment.

2.2.4 Time-inhomogeneous Markov multistate models

The preceding Section 2.2.3 on competing risks generalized the standard survival set-up with one event time T and one event type to modelling different possible event types (the competing risks) that may occur at time T. I.e., competing risks model time until some first event and the type of the first event, but (by definition) potential subsequent events are not modelled. Multistate models allow for modelling both the occurrence of different event types and the occurrence of subsequent events, the latter potentially of different types.

The present section is organized as follows. We first consider some important examples of multistate modelling. In such a model, events are modelled as transitions between different states. Next, we explain that a sequence of events/transitions (i.e., a realization of a multistate process) can be thought of as being realized as a series of nested competing risks experiments. This implies that the estimation techniques of Sections 2.2.1 and 2.2.2 also work in the more complex multistate situation. We then discuss that such a multistate model is time-inhomogeneous Markov and introduce its transition probabilities and transition hazards. Next, the Nelson-Aalen estimator of the cumulative transition hazards is considered, and finally we show how matrix-valued product integration yields the matrix of transition probabilities as a deterministic function of the cumulative transition hazards. Replacing the cumulative transition hazards by their Nelson-Aalen estimators results in the Aalen-Johansen estimator of the transition probabilities.

Examples of a multistate process

In Section 2.1, we explained that we need to keep track of an individual's status over the course of time, which is what a stochastic process $(X_t)_{t\geq 0}$ does. In Section 2.2.3, the realized competing event type X_T naturally arose as the state occupied by the process at event time T. This process point of view becomes indispensable when keeping track of an individual's course through multistate models as depicted in Figures 2.6 and 2.7. In all these models, we

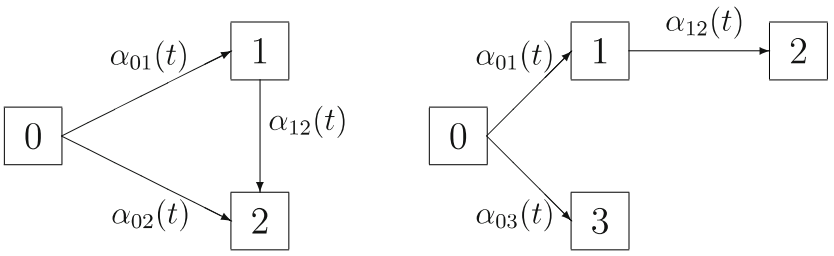

Fig. 2.6. Illness-death models without recovery.

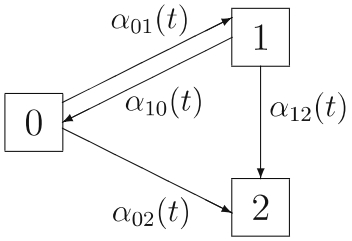

Fig. 2.7. Illness-death model with recovery.

write X_t for the state occupied by the individual at time t.

The model in Figure 2.6 (left) is a so-called illness-death model without recovery. The name of the model stems from the fact that in medical applications state 0 is often interpreted as 'healthy', state 1 is interpreted as 'diseased', and state 2 as 'dead'. The model is 'without recovery', because 'diseased' $\rightarrow$ 'healthy' transitions are not modelled. An individual may either start in the 'healthy' state or in the 'diseased' state (i.e., $X_0 \in \{0, 1\}$).

An individual that starts in the 'healthy' state will have either one or two event times. Say, T is the time the individual leaves its initial state 0. Then $X_t = 0$ for all $t \in [0, T)$ and $X_T \in \{1, 2\}$. If the individual makes a 'healthy' $\rightarrow$ 'dead' transition (i.e., a $0 \rightarrow 2$ transition and $X_T = 2$), there will be no further events for this individual. However, if the individual moves into the 'diseased' state 1 at time T (i.e., $X_T = 1$), there will be a future event time $\tilde{T}$, say, at which the individual moves from 'diseased' to 'dead', $X_{\tilde{T}} = 2$. An individual that starts in the 'diseased' state has only one event time, at which a $1 \rightarrow 2$ transition is made. Note again that the individual is in the 'target state' of

a transition at the event time in question. If we have a $l \to j$ transition at time T, $l \neq j$, then $X_T = j$.

The model in Figure 2.6 (right) is a so-called progressive illness-death model without recovery. The attribute 'progressive' implies that every state has at most one single possible transition into it. There is, in fact, not much difference between the two models of Figure 2.6, still interpreting state 0 as 'healthy' and state 1 as 'diseased'. Both models have the same number of arrows (i.e., transition types), and in both models an individual enters an absorbing state at the time of 'death', which is either state 2 in the non-progressive model or state 2 or 3 in the progressive model. This means that states 2 and 3 of the progressive model have the same interpretation 'death'. The advantage of the progressive model is one of coding: by simply looking at the state entered at time of 'death', one is able to tell whether the individual was 'healthy' or 'diseased' just prior to 'death'. In the non-progressive model, one would also have to look at the state occupied just prior to 'death'.

The model in Figure 2.7 is called an illness-death model with recovery. The difference compared to the models in Figure 2.6 is that now 'diseased' $\to$ 'healthy' transitions (i.e., recoveries) are also possible. Already in the model without recovery, we saw that the number of an individual's event time can be random. However, there were at most two subsequent events. In a model with recovery, there is, at least theoretically, no such maximum number. When actually collecting data, practical restrictions may impose a maximum number. Still, the number of an individual's event times would be random, which further stresses the process way of writing things. As in Section 2.1, an individual's course through model Figure 2.7 is still conveniently written as

$$[0, \infty) \ni t \mapsto X_t,$$

$X_t \in \{0, 1, 2\}$.

Obviously, multistate models can be quite complex. In principle, any finite number of states is admissible, and there may be transitions in both directions between every pair of states. E.g., if states 0 and 1 correspond to two operational levels of a machine, say 'on' and 'off', and state 2 corresponds to 'malfunctioning', the machine may be repaired, such that transitions out of state 2 are also possible. This model specification (i.e., the state space and the possible transitions types between the states) will be directly reflected when using the R packages `mvna`, `etm`, and `mstate`. We also refer to the excellent textbook by Hougaard (2000) who gives a comprehensive account of the different types of multistate models.

Multistate models as a series of nested competing risks experiments

As stated earlier, the multistate models that we consider are realized as successive nested competing risks experiments. For illustration, consider the illness-death model with recovery of Figure 2.7. An individual either starts in state 0

or in state 1 at time origin 0. Consider an individual that starts in state 0; $X_0 = 0$. The course of this individual through the multistate model is realized as follows.

1. Being in state 0, the individual is exposed to 'cause-specific' hazards $\alpha_{01}(t)$ and $\alpha_{02}(t)$. As explained at the end of Section 2.2.3, the individual's waiting time until an event, the time until moving out of state 0, is determined by the 'all-cause' hazard $\alpha_{01}(t) + \alpha_{02}(t)$. The state entered at the time of transition is determined by a binomial experiment and the relative magnitude of the 'cause-specific' hazards $\alpha_{01}(t)$ and $\alpha_{02}(t)$ at the time of transition.
2. The next step depends on the state entered at the first transition time.
   a) If the absorbing state 2 has been entered, there will be no further transition.
   b) If state 1 has been entered, a new competing risks experiment is carried out using the current values of the 'cause-specific' hazards $\alpha_{10}(t)$ and $\alpha_{12}(t)$. Otherwise, the experiment runs analogously to step 1.
3. The next step depends on the state entered at the second transition time. If the absorbing state has been entered, there will be no further transition. If state 0 has been entered, a further competing risks experiment will be carried out.

This series of competing risks experiments will be carried out until absorption. If there is no absorbing state in the model (i.e., backward transitions are feasible out of every state), we will need to keep track of the series of competing risks experiments until observation ends.

As with competing risks, this simulation point of view towards multistate models shows up again and again in the main part of the book. See, in particular, Chapter 8 for an in-depth treatment. A more formal justification of the above algorithm is given in Section 4.4 of Gill and Johansen (1990); see also Theorem II.6.7 of Andersen et al. (1993).

Transition probabilities and transitions hazards of a time-inhomogeneous Markov multistate process

So far, our treatment of multistate models has been a bit lax in that the transitions hazards $\alpha_{lj}(t)$, $l \neq j$, which we indicated at the $l \to j$ arrows in the multistate figures, have been treated as cause-specific hazards of a competing risks model in the above algorithm. Although conceptually correct, a slightly more precise definition is desirable. We also wish to estimate the cumulative transition hazards and derive probability estimates, and we need to state in a more precise manner that multistate models that are realized as successive nested competing risks experiments are time-inhomogeneous Markov.

The Markov property is a key assumption for the estimation techniques discussed below to work with data subject to independent left-truncation and independent right-censoring, respectively. In essence, the Markov property,

which is given a precise form in Equation (2.27) below, means that the future course of an individual depends on the past only via the current time and the state currently occupied by the individual. E.g., the future development of a 'diseased' individual at time t in the models of Figure 2.6 depends on the past only through the time elapsed since time origin (i.e., t and the fact that the individual is currently 'diseased') but not on the time span the individual has already been ill. The analysis of non-Markov models is a quite active research field, on which we briefly comment in Chapter 12.

We begin by defining the matrix of transition probabilities of a Markov process $(X_t)_{t\geq 0}$ with state space $\{0, 1, 2, \ldots, J\}$ as

$$\mathbf{P}(s,t) := \left(\mathrm{P}_{lj}(s,t)\right)_{l,j}, \; l, j \in \{0, 1, 2, \ldots, J\}, \tag{2.26}$$

with transition probabilities

$$\mathrm{P}_{lj}(s,t) := \mathrm{P}(X_t = j \,|\, X_s = l) = \mathrm{P}(X_t = j \,|\, X_s = l, \mathrm{Past}), s \leq t. \tag{2.27}$$

The Markov property $\mathrm{P}(X_t = j \,|\, X_s = l) = \mathrm{P}(X_t = j \,|\, X_s = l, \mathrm{Past})$ intuitively states that past and future of the process are independent given the present at time s. We also note that the Markov process $(X_t)_{t\geq 0}$ is said to be time-inhomogeneous, because the transition probabilities (2.27) depend on the actual time interval $[s, t]$. In contrast, a homogeneous process makes the more restrictive assumption that these probabilities are identical whenever the length of the time interval $d = t - s$ is. The transition probabilities of a homogeneous Markov process only depend on the length of the time interval, but not the interval itself. Readers should note that sometimes homogeneous Markov processes are simply called 'Markov processes', dropping the attribute 'homogeneous'.

Analogous to Definition (2.23) of the cause-specific hazards, we now define the transition hazards of the Markov process

$$\alpha_{lj}(t) \cdot \mathrm{d}t := \mathrm{P}(X_{(t+\mathrm{d}t)-} = j \,|\, X_{t-} = l), \; l, j = 0, \ldots, J, \; l \neq j. \tag{2.28}$$

Note that the Markov property implies that conditioning on $X_{t-} = l$ is tantamount to conditioning on the entire past of the process before t. In words, $\alpha_{lj}(t) \cdot \mathrm{d}t$ is the probability of making an $l \to j$ transition in the very small time interval $\mathrm{d}t$. Intuitively, $\mathrm{d}t$ will be so small that the transition occurs directly from l to j (i.e., without visiting another state in between). Thus, we can think of $\alpha_{lj}(t)$ as momentary forces of transition between states l and j. Formally, we also define

$$\alpha_{ll}(t) = -\sum_{j=0, j\neq l}^{J} \alpha_{lj}(t), \; l = 0, \ldots, J. \tag{2.29}$$

This definition is justified following Equation (2.31) below.

At the beginning of this chapter, we claimed that transition hazards are an intuitively important concept. And, in fact, the $\alpha_{lj}(t)$s of (2.28) are well suited

to contrast time-inhomogeneous Markov models from homogeneous models and from non-Markov models, respectively.

In a homogeneous Markov model, $\mathbf{P}(s,t)$ only depends on $t-s$, but not on the actual time interval. As a consequence, $\alpha_{lj}(t) = \alpha_{lj}(0)$ for all t: a homogeneous model is a parametric model with constant transition hazards. In contrast, the transition hazards (2.28) can essentially be any integrable nonnegative function. Hence, assuming $(X_t)_{t\geq 0}$ to be time-inhomogeneous Markov provides for a much larger nonparametric model.

The restriction implied by the Markov assumption is also well illustrated in terms of the transition hazards. In the illness-death models of Figure 2.6, the 'illness' $\to$ 'death'-hazard is $\alpha_{12}(t)$. It depends on the transition type 'illness' $\to$ 'death' and on the current time t since time origin 0. However, $\alpha_{12}(t)$ does not depend on the entry time $\tilde{t}$, say, into the 'illness'-state 1, $\tilde{t} < t$. In a non-Markov model, the transition hazard would be $\alpha_{12}(\tilde{t}, t)$, which would potentially be different for fixed t but different times $\tilde{t}$ of falling ill.

Nelson-Aalen estimator

As our Definition (2.28) of the transition hazards has been analogous to Definition (2.23) of the cause-specific hazards, it should not come as a surprise that we may estimate the cumulative transition hazards $A_{lj}(t) = \int_0^t \alpha_{lj}(u)\,\mathrm{d}u$ in a manner similar to the cause-specific Nelson-Aalen estimators (2.25). The appropriate Nelson-Aalen estimators are

$$\widehat{A}_{lj}(t) = \sum_{k=1}^{K} \frac{\text{number of observed } l \to j \text{ transitions at } t_k}{\text{number of individuals at risk in state } l \text{ just prior to } t_k}, \quad (2.30)$$

$l, j = 0, \ldots, J$, $l \neq j$, where the summation is over all event times t_k, which are less than or equal to t. As with the cause-specific Nelson-Aalen estimators, we stress a couple of important facts which have been alluded to earlier: the numerator in (2.30) represents increments of a transition-specific counting process. And the risk set in (2.30) includes all individuals who have entered state l before time t_k and who have not yet moved out of state l again or have been censored.

This has three important implications. First, as with the cause-specific Nelson-Aalen estimators, we may code computation of $\widehat{A}_{lj}(t)$ via coding $l \to \tilde{j}$ transitions, $\tilde{j} \neq j$, as censoring events and only count type $l \to j$ transitions as 'actual events'. Second, an individual only contributes to the risk set in state l after entry into the state; movements within a multistate model generate 'internal' left-truncation as explained towards the end of Section 2.2.2. An analysis of $\widehat{A}_{lj}(t)$ must be coded accordingly. Third, every individual in state l and under observation contributes to the risk set alike; there is no further accounting for the individual's entry time into state l. This is a consequence of the Markov assumption. Each of these implications will be directly reflected in R coding.

In fact, one may adopt the view that these three implications are consequences of multistate models being realized as a series of competing risks experiments. It then is via the methodology of Sections 2.2.1 and 2.2.2 that we may analyse the transition hazards of, first, competing risks, and, next, multistate models.

Product integration and the Aalen-Johansen estimator

Finally, we wish to estimate the matrix of transition probabilities $\mathbf{P}(s,t)$. In the simple survival set-up of Section 2.1, we found that we may compute the survival probability as a deterministic function, namely product integration, of the cumulative survival hazard. Replacing the true cumulative hazard by its Nelson-Aalen estimator resulted in the Kaplan-Meier estimator of the survival function. An analogous approach works for estimating $\mathbf{P}(s,t)$.

Analogous to $\mathbf{P}$, we write

$$\mathbf{A}(t) := \left(A_{lj}(t)\right)_{l,j}, \, l,j \in \{0,1,2,\ldots,J\} \tag{2.31}$$

for the matrix of cumulative transition hazards $A_{lj}(t) = \int_0^t \alpha_{lj}(u)\,\mathrm{d}u$. The aim is to show that $\mathbf{P}(s,t)$ can be computed as a continuous matrix-valued product over terms

$$\mathbf{I} + \mathrm{d}\mathbf{A}(u),$$

where u ranges from s to t, where we have written $\mathbf{I}$ for the $(J+1)\times(J+1)$ identity matrix, and where $\mathrm{d}\mathbf{A}(u)$ is defined element wise as

$$\mathrm{d}\left(A_{lj}(u)\right)_{l,j} = \left(\alpha_{lj}(u)\right)_{l,j}\mathrm{d}u,$$

$l,j \in \{0,1,2,\ldots,J\}$. This idea obviously parallels that of Equations (2.5)–(2.7). Also recall from (2.29) that $\mathrm{d}A_{ll}(u) = -\sum_{j=0,j\neq l}^{J}\mathrm{d}A_{lj}(t)$, such that

$$1-\mathrm{d}A_{ll}(u) = 1-\mathrm{P}(X_{(t+\mathrm{d}t)-}\neq l \,|\, X_{t-}=l) = \mathrm{P}(X_{(t+\mathrm{d}t)-}= l \,|\, X_{t-}=l),$$

which explains Definition (2.29), or equivalently why we have to consider a product over terms $\mathbf{I}+\mathrm{d}\mathbf{A}(u)$.

Now consider a time v, $s<v<t$. The Markov property implies that

$$\mathbf{P}(s,t) = \mathbf{P}(s,v)\cdot\mathbf{P}(v,t). \tag{2.32}$$

In order to see that (2.32) holds, consider the (l,j)th entry of $\mathbf{P}(s,t)$:

$$\begin{aligned}\mathrm{P}(X_t=j\,|\,X_s=l) &= \sum_{\tilde{\jmath}=0}^{J}\mathrm{P}(X_v=\tilde{\jmath}\,|\,X_s=l)\cdot\mathrm{P}(X_t=j\,|\,X_v=\tilde{\jmath},X_s=l)\\ &= \sum_{\tilde{\jmath}=0}^{J}\mathrm{P}(X_v=\tilde{\jmath}\,|\,X_s=l)\cdot\mathrm{P}(X_t=j\,|\,X_v=\tilde{\jmath}),\end{aligned}$$

where the last equation holds because of the Markov property, and the right hand side equals the (l, j)th entry of the right hand side of (2.32).

Next, assume that v is close to t such that an approximation such as in Equation (2.7) holds,

$$\sum_{\tilde{\jmath}=0}^{J} \mathrm{P}(X_v = \tilde{\jmath} \,|\, X_s = l) \cdot (\mathbf{1}(\tilde{\jmath} = j) + \Delta A_{\tilde{\jmath}j}(t)),$$

where we have written $\Delta A_{\tilde{\jmath}j}(t)$ for $A_{\tilde{\jmath}j}(t) - A_{\tilde{\jmath}j}(v)$. Doing this recursively for a partition $s = t_0 < t_1 < t_2 < \ldots < t_{K-1} < t_K = t$ of the time interval $[s, t]$, we get the approximation

$$\mathbf{P}(s, t) \approx \prod_{k=1}^{K} \left(\mathbf{I} + \Delta\mathbf{A}(t_k)\right), \tag{2.33}$$

where the (l, j)th element of$\Delta\mathbf{A}(t_k)$ is $A_{lj}(t_k) - A_{lj}(t_{k-1})$. Computing the approximation on the right hand side of (2.33) for ever finer partitions of $[s, t]$ approaches a limit, the matrix-valued product integral $\prod_{u\in(s,t]} (\mathbf{I} + \mathrm{d}\mathbf{A}(u))$. The product integral equals the matrix of transition probabilities,

$$\mathbf{P}(s, t) = \prod_{u\in(s,t]} \!\!\!\!\!\!\prod \left(\mathbf{I} + \mathrm{d}\mathbf{A}(u)\right). \tag{2.34}$$

An estimator of $\mathbf{P}(s, t)$ is now naturally derived by replacing $\mathbf{A}(u)$ with the matrix $\widehat{\mathbf{A}}(u)$ of Nelson-Aalen estimators with the (l, j)th entry $\widehat{A}_{lj}(u)$ as in Equation (2.30) for $l \neq j$ and $\widehat{A}_{ll}(u) := -\sum_{j,j\neq l} \widehat{A}_{lj}(u)$. We also define $\mathrm{d}\widehat{\mathbf{A}}(u)$ as the matrix with entries $\widehat{A}_{lj}(u) - \widehat{A}_{lj}(u-)$ (i.e., the increment of the Nelson-Aalen estimators at time u). This results in the Aalen-Johansen estimator (Aalen and Johansen, 1978),

$$\widehat{\mathbf{P}}(s, t) = \prod_{u\in(s,t]} \!\!\!\!\!\!\prod \left(\mathbf{I} + \mathrm{d}\widehat{\mathbf{A}}(u)\right), \tag{2.35}$$

which is an ordinary, finite matrix product over all event times u in $(s, t]$ and matrices $\mathbf{I} + \mathrm{d}\widehat{\mathbf{A}}(u)$. The Aalen-Johansen estimator is often also called the empirical transition matrix. The estimator and 'empirical' product integration are implemented in the R package `etm`.

We finally note that checking approximation (2.33) in R as we did for the simple survival situation in Section 2.1 is not straightforward. The reason is that closed formulae for $\mathbf{P}(s, t)$ only exist for some special, practically important multistate models; see Section 9.1. In fact, approximation (2.33) provides a numerical tool to compute $\mathbf{P}(s, t)$ in the absence of closed formulae.

2.3 Approximate inference in practice based on large sample results

As with standard survival data, statistical inference for competing risks and multistate models is typically of a nonparametric kind. Asymptotic results are used for approximate inference in practice. E.g., the transition-specific Nelson-Aalen estimator $\widehat{A}_{lj}(t)$ as defined in (2.30) is approximately unbiased in the sense that $\widehat{A}_{lj}(t)$ converges in probability to the true quantity $A_{lj}(t)$. Properly standardized, the distribution of the estimator approaches a normal distribution,

$$\sqrt{n}\left(\widehat{A}_{lj}(t) - A_{lj}(t)\right) \to N(0, \sigma^2_{lj}(t)), \tag{2.36}$$

where n is the number of individuals under study. An (again asymptotically/approximately unbiased) estimator of the asymptotic variance $\sigma^2_{lj}(t)$ is $n \cdot \widehat{\sigma}^2_{lj}(t)$, where

$$\widehat{\sigma}^2_{lj}(t) = \sum_{k=1}^{K} \frac{\text{number of observed } l \to j \text{ transitions at } t_k}{(\text{number of individuals at risk in state } l \text{ just prior to } t_k)^2},$$

where the summation is over all event times t_k, which are less than or equal to t as in (2.30); see also Section 4.1. This can be used, e.g., to construct an approximate 95% confidence interval

$$\widehat{A}_{lj}(t) \pm \widehat{\sigma}_{lj}(t) \cdot 1.96,$$

where $1.96 \approx$ `qnorm(0.975)`, i.e., the 0.975 quantile of the standard normal distribution. (We note, however, that a log-transformed confidence interval should be preferred in small samples; see (4.10).)

A common feature of these approximate procedures is, loosely speaking, that they only hold on the 'observable time interval'. For continuous event times and continuous censoring times, the 'observable time interval' is restricted by the upper limit of the joint support of event time and censoring time. In practice, the 'observable time interval' is considered to be restricted by the largest uncensored event time. Analogous considerations are needed for left-truncated data and for more complex multistate models. A detailed discussion can be found in Examples IV.1.6–IV.1.9 in Andersen et al. (1993). A sufficient condition for the approximate procedures to work is that there is a positive probability of being at risk in a transient state of the multistate model under consideration (Andersen et al., 1993, Equation (4.1.16)). Transient states are those states out of which a transition is possible. E.g., in the competing risks model of Figure 2.5, only the initial state 0 is transient, but in the illness-death models of Figures 2.6 and 2.7, both states 0 and 1 are transient. An individual is said to be at risk in a transient state l just prior to time t, if the individual is in state l and under observation at $t-$. Only such an individual may be observed to make a transition out of state l at time t.

Requiring a positive probability of a non-empty risk set has different implications depending on which quantity is being estimated. If the aim is to estimate the cumulative transition hazard between states l and j of a certain multistate model, $l \neq j$, the requirement only affects the risk set Y_l. Estimation of probabilities, however, in general depends on the estimation of all cumulative hazards because of relations (2.34) and (2.35).

For illustration, we briefly comment on a basic implication in the simple competing risks model. Often, there is interest in the failure type probabilities $\mathrm{P}(X_T = j)$, where j is one of the competing event states $1, 2, \ldots, J$ as in Figure 2.5. E.g., Mackenbach et al. (1999) investigated such 'prevalences' of causes of death in the Netherlands. $\mathrm{P}(X_T = j)$ is the limit of the so-called cumulative incidence function,

$$\mathrm{P}(X_T = j) = \lim_{t \to \infty} \mathrm{P}(T \leq t, X_T = j).$$

Estimation of $\mathrm{P}(T \leq t, X_T = j)$ is a special application of the Aalen-Johansen estimator (2.35). Estimation of its limit $\mathrm{P}(X_T = j)$ is simple, if the data are complete. The Aalen-Johansen estimator evaluated at the largest time point then simply equals the so-called 'crude rates', the number of type j events (at any time), divided by the sample size n. However, $\mathrm{P}(X_T = j)$ will not be nonparametrically estimable with most right-censored data. The reason for this is that censoring typically restricts the 'observable time interval' such that one will not be able to observe the limit of the cumulative incidence function. See also our discussion following (4.21).

It is worthwhile to note that approximate unbiasedness and approximate normality hold *uniformly* on what we have loosely called an 'observable time interval'. For weak convergence, this requires a theory of convergence of probability measures on a space of functions rather than the well-known concept of weak convergence of distribution functions. This functional point of view is, e.g., relevant when moving from the Nelson-Aalen estimator $\widehat{\mathbf{A}}$ to the Aalen-Johansen estimator $\widehat{\mathbf{P}}(s,t)$; see (2.35). This is so, because $\widehat{\mathbf{P}}(s,t)$ is a function of *all* previous Nelson-Aalen estimates between s and t, i.e., all $\widehat{\mathbf{A}}(u)$, $u \in (s,t]$.

The mathematics of such a convergence theory are formidable and beyond the technical level of this book. We are content with the fact that asymptotic unbiasedness and asymptotic normality can be formulated rigorously and sufficiently general for the applications in this book. The generally interested reader is referred to the excellent books by Billingsley (1968), Andersen et al. (1993), and van der Vaart and Wellner (1996). In particular, Billingsley's introduction gives a very readable account of why a functional point of view is useful. On the other hand, his Section 3.18 concisely explains why obtaining *uniform* results is difficult. These difficulties have been solved using the modern theory of empirical processes, of which van der Vaart and Wellner give a definite account. Finally, Andersen et al. give a dense but thorough description of asymptotic theory for event history analysis; see, in particular, their Section II.8.

We must, however, still mention the functional delta method as an important tool from the theory of empirical processes. The usual delta method starts with standard, i.e., pointwise convergence as in (2.36). Considering a differentiable transformation $\phi : \mathbb{R} \to \mathbb{R}$ with derivative ϕ', the delta method implies that

$$\sqrt{n}\left(\phi(\widehat{A}_{lj})(t) - \phi(A_{lj})(t)\right) \to \phi'(A_{lj}(t))N(0, \sigma^2_{lj}(t)), \tag{2.37}$$

and that the left-hand side of (2.37) is asymptotically equivalent to $\phi'(A_{lj}(t)) \cdot \sqrt{n}\left(\widehat{A}_{lj}(t) - A_{lj}(t)\right)$, i.e., the difference of the asymptotically equivalent terms converges to zero in probability. We use the ordinary delta method for obtaining pointwise confidence intervals of the Nelson-Aalen estimator based on a log-transformation; see (4.10). A generalization of the delta method to p vectors and transformations $\phi : \mathbb{R}^p \to \mathbb{R}^q$ is immediately available, but what is really needed is a *functional* delta method. This does exist (Gill, 1989), but is again beyond the technical level of this book. We note, however, that the functional delta method works for product integration as in Equations (2.34) and (2.35). This further emphasizes the key roles played by both the Nelson-Aalen estimator and the product integral. We mention two further important consequences. The functional delta method preserves asymptotic normality and it justifies using bootstrap resampling as discussed in Appendix A; see Gill (1989) and van der Vaart and Wellner (1996). This is helpful in situations where variance estimators are analytically hardly tractable. The variance may then be estimated based on the bootstrap and confidence intervals may again be constructed based on approximate normality.

2.4 Exercises

1. Show that an event time T with hazard $\alpha(t)$ has distribution function $P(T \leq t) = 1 - \exp(-\int_0^t \alpha(u)\,\mathrm{d}u)$.
2. Write a function `A.gompertz` for the cumulative hazard when the survival time distribution follows a Gompertz distribution with shape parameter $\lambda = 1$ and scale parameter $\gamma = 2$.
   Under the Gompertz distribution, the hazard is
   $$\alpha(t) = \lambda \exp(t/\gamma).$$
   Using the `prodint` function from Section 2.1.1, approximate $S(1) = \mathrm{P}(T > 1)$ and compare it to the true value.
3. Simulate 100 individuals with survival times following a Gompertz distribution with parameters as in Exercise 2. Also simulate independent censoring times following a uniform distribution in order to obtain approximately 30% of censored observations. A function to generate Gompertz random variables can be found in the R package `eha`.

Estimate $S(1)$ both using `prodint` and the `survfit` function.
Using the output of `survfit`, compute the Nelson-Aalen estimator of the cumulative hazard and check whether it is close to the true cumulative hazard function.

4. Plot the counting process of observed events and its compensator as in Figure 2.3. Check that the compensator is only almost a smooth line by displaying some 'edges'.
5. Redo the analysis of Exercise 3, but this time with at least 50% censoring.
6. Reuse the simulated data set from Exercise 3 and additionally simulate independent left-truncation times which follow a Weibull distribution. Choose the parameters such that approximately 70% of the simulated individuals are actually included in the study. Check that the Nelson-Aalen estimator 'works' in the presence of left-truncation and right-censoring.
7. *Competing risks*: Simulate competing risks data for 200 individuals with constant cause-specific hazards $\alpha_{01}(t) = 0.5$ and $\alpha_{02}(t) = 0.9$. Independent right-censoring times follow a uniform distribution with parameter chosen to give approximately 20% of censored observations.
   Compute the cause-specific Nelson-Aalen estimators.
8. Definition (2.22) of the cause-specific hazards implies that

$$P(T \leq t, X_T = 1) = \int_0^t P(T \geq u-)\alpha_{01}(u)\,\mathrm{d}u.$$

   Show that

$$P(T \leq t, X_T = 1) \leq 1 - \exp(-\int_0^t \alpha_{01}(u)\,\mathrm{d}u).$$

   One minus the right hand side of the previous equation is sometimes called the 'cause-specific survivor function'. The right hand side of the equation can be estimated using one minus a Kaplan-Meier-type estimator, but it lacks a proper probability interpretation. Check that this estimator overestimates $P(T \leq t, X_T = 1)$ using the simulated competing risks data.
9. *Multistate models*: Simulate data from an illness-death model without recovery. All individuals are assumed to start in an initial state 0, 'healthy'. Hazards out of the initial state are as in Exercise 7. For individuals who reach state 1, simulate new event times $\tilde{T}$ with constant hazard $\alpha_{12}(t) = 0.8$. $T + \tilde{T}$ will then be the time of entry into state 2, 'death', for individuals who have moved through the 'illness' state 1.
   Estimate the cumulative transition hazards for the following scenarios.
   a) Complete data.
   b) Randomly right-censored data: Draw uniformly distributed censoring times C such that approximately 20% of the observations are censored in the initial state.

c) State-dependent censoring: Assume that individuals who are *observed* to move into state 1 are subject to censoring times $\tilde{C}$ which follow a uniform distribution that is different from the distribution of C.
d) Repeat the previous analyses, but additionally introduce random left-truncation, with left-truncation times stemming from a gamma distribution with parameters chosen to let approximately 90% of the individuals enter the study.

10. *Time-inhomogeneous Markov property*: Show that a competing risks process fulfills the Markov property. When is an illness-death model without recovery Markov?

http://www.springer.com/978-1-4614-2035-4

Competing Risks and Multistate Models with R
Beyersmann, J.; Allignol, A.; Schumacher, M.
2012, XI, 245 p. 49 illus.,
ISBN: 978-1-4614-2035-4