Survival analysis notes

Rob Trangucci

January 15, 2025

Contents

Chapter 1

Notation

Notation	Description
C_i	Random variable representing the time to censoring
$T_i = \min(X_i, C_i)$	Observable event time
$\delta_i = \mathbb{1}(T_i = X_i)$	Indicator variable equal to one if event time is a failure time
$P_\theta(X \leq t)$	Distribution function of X indexed by parameters θ
$S_X(t; \theta)$	Survival function for random variable X evaluated at t, parameters θ
$f_X(t; \theta)$	Density function for random variable X evaluated at t, parameters θ
$\lambda_X(t; \theta)$	Hazard function for random variable X evaluated at t, parameters θ

Table 1.1: List of notation used throughout the notes

Chapter 2

Introduction

This introduction is based in part on Klein, Moeschberger, et al. 2003, and in part on Aalen et al. 2008 plus Fleming and Harrington 2005.

Survival analysis is the modeling and analysis of time-to-event data; this means we will be studying how to model **nonnegative** random variables (time will always be measured in such a way so that the observations are nonnegative). Think about a clinical trial for a new COVID vaccine and how you might model the length of time between study entry and infection in each arm of the trial. Let X_i be the time from trial entry to infection for the i-th participant. These sorts of trials are typically run until a prespecified number of people have become infected. Let n be the total number of participants in the trial and let r be the prespecified number of infections. Let T_i be the observed infection time for the i-th participant. This means that for r participants, $T_i = X_i$, but for $n - r$ participants we know only that the time-to-infection is larger than the observed time. Let C_i denote the time from study entry for participant i to study end. Then $T_i = \min(X_i, C_i)$, and let $\delta_i = \mathbb{1}(T_i = X_i)$. The density of T_i is related to the joint probability for X_i and C_i, which is indexed by a possibly infinite dimensional parameter θ: $P_\theta(X_i > t, C_i > c)$. When $\delta_i = 1$, and $T_i = X_i$, the likelihood of the observation is

$$\left(-\frac{\partial}{\partial u} P_\theta(X_i > u, C_i > t)\right)\Big|_{u=t},$$

while the likelihood for $\delta_i = 0$ is

$$\left(-\frac{\partial}{\partial u} P_\theta(X_i > t, C_i > u)\right)\Big|_{u=t},$$

Then $T_i = C_i$ for the other $n - r$ participants. Under the null hypothesis that the vaccine has no effect, the population distribution function for all n participants for X_i, C_i is $P_\theta(X_1 > x, C_1 > c)$ (i.e. the distribution for survival times in the treatment group and the placebo

group is the same). Then the joint density for the observed infection times is as follows:

$$f_{T_1,\ldots,T_n}(t_1,\ldots,t_n;\theta) = n!\prod_{i=1}^{r}\left(-\frac{\partial}{\partial u}P_\theta(X_1 > u, C_1 > t_{(i)})\right)\Big|_{u=t_{(i)}}\prod_{i=r+1}^{n}\left(\left(-\frac{\partial}{\partial u}P_\theta(X_1 > t_{(i)}, C_1 > u)\right)\Big|_{u=t_{(i)}}\right),$$

where $t_{(i)}$ is the i-th order statistic of the set $\{t_1,\ldots,t_n\}$. Note that this is different from most other data analysis where missing observations are not expected to occur with much frequency. On the contrary, in survival analysis, missingness, both *truncation* and *censoring* are expected to occur with nearly every dataset, so much of our time will be spent ensuring our methods work when data arise with these peculiarities.

2.1 Independent censoring

Now suppose that $X_1 \perp\!\!\!\perp C_1$, and that θ partitions into η and ϕ, such that

$$P_\theta(X_1 > x, C_1 > c) = P_\eta(X_1 > x)P_\phi(C_1 > c).$$

Then we can rewrite the joint observational density for T_i as:

$$\begin{aligned} f_{T_1,\ldots,T_n}(t_1,\ldots,t_n;\theta) = n!&\left(\prod_{i=1}^{r} f_{X_1}(t_{(i)};\eta)\right)\prod_{i=r+1}^{n} P_\eta(X_1 > t_{(i)}) \\ &\times\left(\prod_{i=1}^{r} P_\phi(C_1 > t_{(i)})\right)\prod_{i=r+1}^{n} f_C(t_{(i)};\phi). \end{aligned}$$

If we are only interested about inference about η, the parameters that govern the distribution of the true time-to-infection random variables, we can ignore the the distribution for the censoring random variables C_1, and maximize the likelihood because, in η:

$$f_{T_1,\ldots,T_n}(t_1,\ldots,t_n;\eta) \propto \left(\prod_{i=1}^{r} f_{X_1}(t_{(i)};\eta)\right)\prod_{i=r+1}^{n} P_\eta(X_1 > t_{(i)})$$

We will talk in more detail about censoring in the coming lectures.

2.2 Mean time to failure

Aalen et al. 2008 notes that we cannot even compute a simple mean in this situation, so something like a t-test will be useless. As an aside, let's try to compute a mean from the data above. Let $\bar{T} = \frac{1}{n}\sum_{i=1}^{n} T_i$. We can show that $\lim_{n\to\infty}\bar{T} \le \mathbb{E}[X_i]$ with probability 1.

Proof. Let $T_i = X_i\mathbb{1}(X_i \le C_i) + C_i\mathbb{1}(X_i > C_i)$. Then by the SLLN $\bar{T} \overset{\text{a.s.}}{\to} \mathbb{E}[T_i]$.

$$\begin{aligned} \mathbb{E}[T_i] &= \mathbb{E}[X_i\mathbb{1}(X_i \le C_i)] + \mathbb{E}[C_i\mathbb{1}(X_i > C_i)] \\ &\le \mathbb{E}[X_i\mathbb{1}(X_i \le C_i)] + \mathbb{E}[X_i\mathbb{1}(X_i > C_i)] = \mathbb{E}[X_i] \end{aligned}$$

□

2.3 Survival function

How can we compute the mean time to infection then? One way to estimate the mean time to infection is to first estimate the function $S_{X_i}(t;\theta) = P_\theta(X_i > t)$, which is also known as the *survival function*. Recall this fact about non-negative random variables $X_i \geq 0$ w.p. 1:

$$\mathbb{E}[X_i] = \int_0^\infty P_\theta(X_i > t)dt$$

This follows from an application of Fubini's theorem applied to the integral:

$$\begin{aligned}
\mathbb{E}[X_i] &= \int_0^\infty u dP_{X_i}(u;\theta) \\
&= \int_0^\infty \int_0^\infty \mathbb{1}(0 \leq t \leq u)\, dt\, dP_{X_i}(u;\theta) \\
&= \int_0^\infty \int_0^\infty \mathbb{1}(0 \leq t \leq u)\, dP_{X_i}(u;\theta)dt \\
&= \int_0^\infty P_\theta(X_i > t)dt
\end{aligned}$$

2.3.1 Properties of the survival function

Let $F_{X_i}(t;\theta) = P_\theta(X_i \leq t)$. Then because the survival function is defined as $S_{X_i}(t;\theta) = 1 - F_{X_i}(t;\theta)$ (also known as the complementary CDF) the survival function inherits its properties from the CDF. The survival function:

1. $S_{X_i}(t;\theta)$ is a nonincreasing function
2. $S_{X_i}(0;\theta) = 1$
3. $\lim_{t\to\infty} S_{X_i}(t;\theta) = 0$
4. Has lefthand limits:
$$\lim_{s \nearrow t} S_{X_i}(s;\theta) = S_{X_i}(t-;\theta).$$
5. Is right continuous:
$$\lim_{s \searrow t} S_{X_i}(s;\theta) = S_{X_i}(t;\theta).$$

An example of a discrete survival function is shown in Figure 2.1.

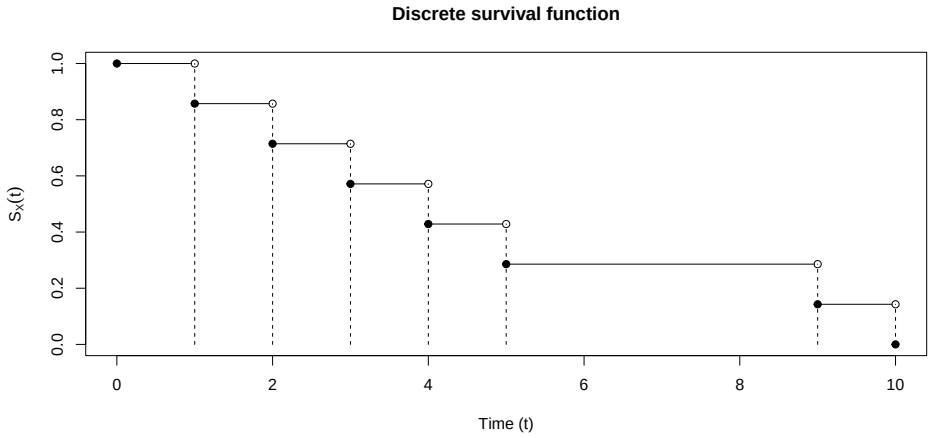

Figure 2.1: Example plot of a survival function for a discrete survival time, bounded between $[0, 10]$

2.4 Hazard function

Another way to characterize the random variable X_i is the *hazard function*, which is typically denoted as $\lambda(t)$ or $h(t)$ and is defined as

$$\begin{aligned}\lambda_{X_i}(t) &= \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathbb{P}_\theta \left(t \leq X_i < t + \Delta t \mid X_i \geq t\right) \\ &= \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \frac{\mathbb{P}_\theta \left(t \leq X_i < t + \Delta t\right)}{\mathbb{P}_\theta \left(X_i \geq t\right)}\end{aligned}$$

First, note that we can define $\mathbb{P}_\theta \left(X_i \geq t\right)$ in terms of the survival function as:

$$\mathbb{P}_\theta \left(X_i \geq t\right) = \lim_{s \nearrow t} S_{X_i}(s; \theta).$$

Using the notation introduced in Section 2.3.1, we can write this as

$$\mathbb{P}_\theta \left(X_i \geq t\right) = S_{X_i}(t-; \theta).$$

Of course, when X_i is absolutely continuous,$S_{X_i}(t-; \theta) = S_{X_i}(t; \theta)$, but when X_i is discrete, or mixed discrete and continuous, as noted above, it is not true in general that the survival function is left-continuous.

A few things to note about $\lambda_{X_i}(t; \theta)$: when X_i is an absolutely continuous random variable, which occurs when we're considering survival in continuous time, we can write this in terms of the probability density function $f_{X_i}(t; \theta)$ and the cumulative distribution function $F_{X_i}(t; \theta)$:

$$\begin{aligned}\lambda_{X_i}(t) &= \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \frac{\mathbb{P}_\theta \left(t \leq X_i < t + \Delta t\right)}{\mathbb{P}_\theta \left(X_i \geq t\right)} \\ &= \lim_{\Delta t \searrow 0} \frac{F_{X_i}(t + \Delta t; \theta) - F_{X_i}(t; \theta)}{\Delta t} \times \frac{1}{1 - F_{X_i}(t; \theta)} \\ &= \frac{f_{X_i}(t; \theta)}{1 - F_{X_i}(t; \theta)}.\end{aligned}$$

Let's examine how the survival function and the hazard function fit together.

$$\lambda_{X_i}(t) = \frac{f_{X_i}(t; \theta)}{S_{X_i}(t-; \theta)}.$$

Note that we can write the hazard function in terms of the survival function instead of the density, when X_i is absolutely continuous:

$$\begin{aligned}\lambda_{X_i}(t) &= \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \frac{\mathbb{P}_\theta \left(t \leq X_i < t + \Delta t\right)}{\mathbb{P}_\theta \left(X_i \geq t\right)} \\ &= \lim_{\Delta t \searrow 0} \frac{S_{X_i}(t; \theta) - S_{X_i}(t + \Delta t; \theta)}{\Delta t} \times \frac{1}{S_{X_i}(t; \theta)} \\ &= -\frac{d}{dt} S_{X_i}(t; \theta) / S_{X_i}(t; \theta).\end{aligned}$$

This implies that

$$\lambda_{X_i}(t) = -\frac{d}{dt}\log S_{X_i}(t;\theta).$$

If we integrate both sides, we get another important identity in survival analysis:

$$\int_0^u \frac{d}{dt}\log S_{X_i}(t;\theta)dt = -\int_0^u \lambda_{X_i}(t)dt \tag{2.1}$$

$$\log S_{X_i}(u;\theta) - \log S_{X_i}(0;\theta) = -\int_0^u \lambda_{X_i}(t)dt \quad \text{note } S_{X_i}(0;\theta) = 1 \tag{2.2}$$

$$S_{X_i}(u;\theta) = \exp\left(-\int_0^u \lambda_{X_i}(t)dt\right) \tag{2.3}$$

2.4.1 Properties of the hazard function

The relationship $S_{X_i}(u;\theta) = \exp\left(-\int_0^u \lambda_{X_i}(t)dt\right)$ and the properties of the survival function reveal the following facts about the hazard function and highlight its differences with a probability density.

1. $\lim_{t\to\infty} S_{X_i}(t;\theta) = 0$ implies that $\lim_{t\to\infty}\int_0^t \lambda_X(u)du = \infty$
2. Given that $S_{X_i}(t;\theta)$ is a nonincreasing function, $\lambda_X(t) \geq 0$ for all t.

So unlike a probability density function, $\lambda_X(t)$ isn't integrable over the support of the random variable.

2.5 Density function for survival time

Given that we have $S_{X_i}(t;\theta)$ and $\lambda(t) = \frac{f_{X_i}(t;\theta)}{S_{X_i}(t-;\theta)}$, we can recover the density, $f_{X_i}(t;\theta)$ easily:

$$f_{X_i}(t;\theta) = \lambda_{X_i}(t)S_{X_i}(t;\theta)$$

2.6 Cumulative hazard function

One final important quantity that describes a survival distribution is that of *cumulative hazard*, which we'll denote as $\Lambda_{X_i}(t)$, though it is also denoted as $H(t)$ in Klein, Moeschberger, et al. 2003. This is defined as you might expect:

$$\Lambda_{X_i}(t) = \int_0^t \lambda_{X_i}(u)du.$$

It has the important property that for any absolutely continuous failure time X_i with a given cumulative hazard function, the random variable $Y_i = \Lambda_{X_i}(X_i)$ is exponentially distributed

with rate 1. The derivation is straightforward. Remember that $P(X_i > t) = \exp\left(-\Lambda_{X_i}(t)\right)$

$$\begin{aligned} P(\Lambda_{X_i}(X_i) > t) &= P(X_i > \Lambda_{X_i}^{-1}(t)) \\ &= \exp\left(-\Lambda_{X_i}(\Lambda_{X_i}^{-1}(t))\right) \\ &= \exp\left(-t\right) \end{aligned}$$

2.7 Discrete survival time

We've been working with continuous survival times until now. If X_i is a discrete random variable with support on $\{t_1, t_2, \dots\}$, we lose some of the tidyness of the previous derivations. We can define the distribution of X_i in terms of the survival function, $P_\theta(X_i > t)$. First let $p_j = P_\theta(X_i = t_j)$, so

$$S_{X_i}(t;\, \theta) = P_\theta(X_i > t) = \sum_{j|t_j > t} p_j$$

We can also define the hazard function for a discrete random variable:

$$\lambda_{X_i}(t_j) = \frac{p_j}{S_{X_i}(t_{j-1};\, \theta)} = \frac{p_j}{p_j + p_{j+1} + \dots}$$

Note that $p_j = S_{X_i}(t_{j-1};\, \theta) - S_{X_i}(t_j;\, \theta)$, then

$$\lambda_{X_i}(t_j) = 1 - \frac{S_{X_i}(t_j;\, \theta)}{S_{X_i}(t_{j-1};\, \theta)}.$$

If we let $t_0 = 0$ then $S_{X_i}(t_0;\, \theta) = 1$. This allows us to write the survival function in a sort of telescoping product:

$$\begin{aligned} P_\theta(X_i > t_j) &= P_\theta(X_i > t_0)\frac{P_\theta(X_i > t_1)}{P_\theta(X_i > t_0)}\frac{P_\theta(X_i > t_2)}{P_\theta(X_i > t_1)} \cdots \frac{P_\theta(X_i > t_j)}{P_\theta(X_i > t_{j-1})} \\ &= 1\frac{S_{X_i}(t_1;\, \theta)}{S_{X_i}(t_0;\, \theta)}\frac{S_{X_i}(t_2;\, \theta)}{S_{X_i}(t_1;\, \theta)} \cdots \frac{S_{X_i}(t_j;\, \theta)}{S_{X_i}(t_{j-1};\, \theta)} \end{aligned}$$

This yields another way to write $S_{X_i}(t;\, \theta)$:

$$S_{X_i}(t;\, \theta) = \prod_{j|t_j \le t} (1 - \lambda_{X_i}(t_j)). \tag{2.4}$$

It turns out that we can write the survival function for continuous random variables in the same way.

2.7.1 Connection between discrete and continuous survival functions

Recall the definition of the hazard function:

$$\lambda_{X_i}(t) = \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathbb{P}_\theta \left(t \leq X < t + \Delta t \mid X \geq t \right)$$

Note that $\lambda_{X_i}(t)\,\Delta t$ is approximately $\mathbb{P}_\theta \left(t \leq X < t + \Delta t \mid X \geq t \right)$. Let $\mathcal{T}$ be a partition of $(0, \infty)$ with partition size Δt, $t_0 = 0$:

$$\mathcal{T} = \bigcup_{j=0}^{\infty} [t_j, t_j + \Delta t).$$

Then we can use Equation (2.4) to represent the survival function:

$$S_{X_i}(t;\, \theta) = \prod_{j | t_j + \Delta t \leq t} (1 - \lambda_{X_i}(t_j) \Delta t). \tag{2.5}$$

We can show that as the partition of the time domain gets finer and finer, we will recover $S_{X_i}(t;\, \theta) = \exp(-\int_0^t \lambda_{X_i}(u) du)$

$$S_{X_i}(t;\, \theta) = \prod_{j \in \mathcal{T} | t_j + \Delta t \leq t} (1 - \lambda_{X_i}(t_j) \Delta t) \tag{2.6}$$

$$\log S_{X_i}(t;\, \theta) = \sum_{j \in \mathcal{T} | t_j + \Delta t \leq t} \log(1 - \lambda_{X_i}(t_j) \Delta t) \tag{2.7}$$

We use the Taylor expansion of $\log(1 - \lambda_{X_i}(t_j)\Delta t)$ for small $\lambda_{X_i}(t_j)\Delta t$, assuming that $\lambda_{X_i}(t)$ is sufficiently well-behaved for all t.

$$\log(1 - \lambda_{X_i}(t_j) \Delta t) \approxeq -\lambda_{X_i}(t_j) \Delta t.$$

Then

$$\log S_{X_i}(t;\, \theta) \approxeq \sum_{j \in \mathcal{T} | t_j + \Delta t \leq t} -\lambda_{X_i}(t_j) \Delta t \tag{2.8}$$

As

$$\lim_{\Delta t \searrow 0} \sum_{j \in \mathcal{T} | t_j + \Delta t \leq t} -\lambda_{X_i}(t_j) \Delta t = -\int_0^t \lambda_{X_i}(u) du.$$

So, $S_{X_i}(t;\, \theta) = \exp(-\int_0^t \lambda_{X_i}(u) du)$, or

$$S_{X_i}(t;\, \theta) = \exp(-\lambda_{X_i}(t)) \tag{2.9}$$

2.8 Mean residual lifetime

We also might be interested in the *mean residual lifetime* (mrl for short), or the expected lifetime given survival up to a certain point:

$$\mathbb{E}\left[X_i - x \mid X_i > x\right].$$

We can compute this for an absolutely continuous random variable by using the survival function:

$$\frac{\int_x^\infty (u-x) f_{X_i}(u;\theta)du}{S_{X_i}(x;\theta)} = \frac{\int_x^\infty S_{X_i}(u;\theta)du}{S_{X_i}(x;\theta)}$$

To derive the mrl in terms of the survival function, note that we can use Fubini again on the numerator (Exercise 1), or we can use integration by parts:

$$\begin{aligned}\int_x^\infty (u-x) f_{X_i}(u)du &= -\int_x^\infty (u-x)\frac{d}{du}S_{X_i}(u)du \\ &= -(u-x)S_{X_i}(u)\big|_{u=x}^\infty + \int_x^\infty S_{X_i}(u)du\end{aligned}$$

and use the fact that $\lim_{u\to\infty} S_{X_i}(u) = 0$. We also need the following:

$$\lim_{u\to\infty} uP(X_i > u) = 0. \tag{2.10}$$

This is a pretty weak condition, random variables with second moments satisfy this condition (Exercise 2), as do random variables with only first moments. It turns out that under this condition we'll have a weak law of large numbers (see §7.1 in Resnick 2019).

Suppose we assume that $\mathbb{E}[X] \leq \infty$. Then we can write:

$$\mathbb{E}[X] = \mathbb{E}[X\mathbb{1}(X \leq n)] + \mathbb{E}[X\mathbb{1}(X > n)]$$

Note that if we define $X_n = X\mathbb{1}(X \leq n)$ then

$$X_1(\omega) \leq X_2(\omega) \leq \cdots \leq X_k(\omega) \leq \ldots.$$

By the Monotone Convergence Theorem (MCT), $\mathbb{E}[X_n] \to \mathbb{E}[X]$. Then

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X_n] + \mathbb{E}[X\mathbb{1}(X > n)] \\ &\geq \mathbb{E}[X_n] + \mathbb{E}[n\mathbb{1}(X > n)] \\ &= \mathbb{E}[X_n] + nP(X > n)\end{aligned}$$

This leads to the system of inequqalities:

$$\mathbb{E}[X] - \mathbb{E}[X_n] \geq nP(X > n) \geq 0.$$

By the MCT $\mathbb{E}[X] - \mathbb{E}[X_n] \to 0$ so

$$\lim_{n\to\infty} nP(X_i > n) = 0.$$

However, there are random variables for which $\mathbb{E}[X_i]$ does not exist, but do satisfy Equation (2.10) (see the end of §7.1 in Resnick 2019).

2.9 Examples

The first example we'll run through is for an exponentially distributed survival time:

$$X_i \overset{\text{iid}}{\sim} \text{Exp}(\lambda).$$

The survival function is $S_X(t) = e^{-\lambda t}$. We can read off from this that $\Lambda(t) = \lambda t$. What's the hazard function? Let's plot the hazard function. What does this imply about the exponential distribution (memorylessness)? The mean lifetime is $\frac{1}{\lambda}$. The mean residual lifetime is:

$$\begin{aligned}\frac{\int_t^\infty e^{-\lambda u} du}{e^{-\lambda t}} &= \frac{1}{\lambda}\frac{e^{-\lambda t} du}{e^{-\lambda t}} \\ &= \frac{1}{\lambda}.\end{aligned}$$

This is a consequence of the memoryless property of the exponential distribution.

Another parametric distribution for survival times is the Weibull.

$$X_i \overset{\text{iid}}{\sim} \text{Weibull}(\gamma, \alpha).$$

The survival function:

$$S_X(t) = \exp(-\gamma t^\alpha).$$

Again, we have that $\Lambda(t) = \gamma t^\alpha$, so we can take the derivative with respect to t to get the hazard:

$$\lambda(t) = \gamma\alpha t^{\alpha-1}.$$

This is more flexible than the exponential distribution, though note that for $\alpha = 1$, $X_i \sim$ Exponential(γ), so the Weibull family contains the exponential family as a special case. The α parameter allows for the hazard rate to have more flexibility than the exponential. If $\alpha > 1$, the hazard rate is increasing in t. This corresponds to an aging process, whereby the longer something has survived, the higher the rate of failure. If $\alpha < 1$, the hazard rate is decreasing in t. This might correspond to something like the hazard for SIDS, which is quite high for

children before 1 year old, but drops off rapidly after 1. Let's compute the mean lifetime, $\mathbb{E}[X] = \int_0^\infty S_X(t)dt$, using a v-sub, $v = t^\alpha$, so $v^{\frac{1}{\alpha}} = t \to \frac{1}{\alpha}v^{\frac{1}{\alpha}-1}dv = dt$:

$$\begin{aligned}\int_0^\infty \exp(-\gamma t^\alpha)dt &= \frac{1}{\alpha}\int_0^\infty v^{\frac{1}{\alpha}-1}\exp(-\gamma v)dv \\ &= \frac{1}{\alpha}\frac{1}{\gamma^{\frac{1}{\alpha}}}\Gamma(\frac{1}{\alpha}) \\ &= \frac{\Gamma(\frac{1}{\alpha}+1)}{\gamma^{\frac{1}{\alpha}}}\end{aligned}$$

The mean residual lifetime is a bit more involved. Let $v = \gamma u^\alpha$ so $\left(\frac{v}{\gamma}\right)^{1/\alpha} = u \to \gamma^{-1/\alpha}\frac{1}{\alpha}v^{\frac{1}{\alpha}-1}dv = du$:

$$\begin{aligned}\int_t^\infty \exp(-\gamma u^\alpha)du &= \gamma^{-1/\alpha}\frac{1}{\alpha}\int_{\gamma t^\alpha}^\infty v^{\frac{1}{\alpha}-1}\exp(-v)dv \\ &= \gamma^{-1/\alpha}\frac{1}{\alpha}\Gamma(\frac{1}{\alpha}, \gamma t^\alpha),\end{aligned}$$

where $\Gamma(\frac{1}{\alpha}, \gamma t^\alpha)$ is the upper incomplete Gamma function.

Chapter 3

Censoring and truncation

Now let's delve into more detail about censoring, and how the likelihood can be built up from the hazard function and the survival function. Klein, Moeschberger, et al. 2003 define censoring as imprecise knowledge about an event time. If we observe a failure or an event exactly, the observation is not censored, but if we know only that an observation occurred within a range of values, we say the observation is censored. Let X_i, as usual, be our failure time, which is not completely observed. Instead if:

- $X_i \in [U, \infty)$, the observation is *right censored*
- $X_i \in [0, V)$, the observation is *left censored*
- $X_i \in [U, V)$, the observation is *interval censored*

3.1 Right censoring

Right censoring occurs when a survival time is known to be larger than a given value. This is the most common censoring scenario in survival analysis.

Recall our definition in Chapter 2:

- Let X_i be the time to failure, or time to event for individual i.
- Let C_i be the time to censoring. It may be helpful to think about C_i as the time to investigator measurement.
- Let $\delta_i = \mathbb{1}(X_i \leq C_i)$.
- Let $T_i = \min(X_i, C_i)$.

Given our definitions in Section 3.1, when an observation is censored, or when a measurement is taken of the survival time before the event has happened, $\delta_i = 0$ and $T_i = C_i$.

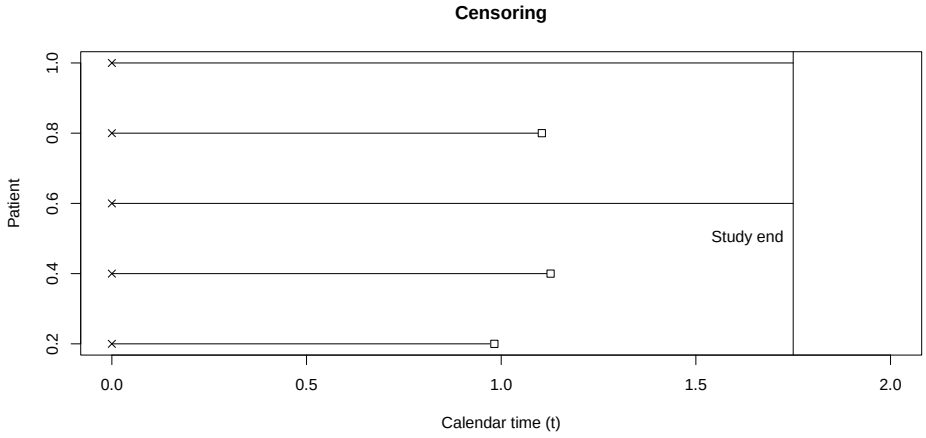

Figure 3.1: Example of Type I censoring.

3.1.1 Type I censoring

The simplest censoring scenario is one in which all individuals have the same, nonrandom censoring time. Imagine a study is designed to follow 5 startups that are spun out of a tech incubator to study how long it takes a company to land its first contract. This information will be used for designing investments 2 years from the study date, so the study has a length of 1.75 years. We can say that all observations will have to have occurred, or not, by 1.75 years.

Figure 3.1 shows a potential result of the study, where 2 out of the 5 companies have not landed a contract. In this case,

- For all individuals such that $\delta_i = 0 \implies X_i > C$
- $\delta_i = 1 \implies T_i = X_i$.

3.1.2 Generalized type I censoring

A more general scenario, which is closer to most examples in clinical trials, is when each individual has a different study entry time and the investigator has a preset study end time. This is called generalized Type I censoring. These study entry times are typically assumed to be independent of the survival time. This is shown in Figure 3.2. When study entry is independent from survival time, the analysis proceeds as shown in Figure 3.3. For generalized type I censoring,

- For all individuals such that $\delta_i = 0 \implies X_i > C_i$

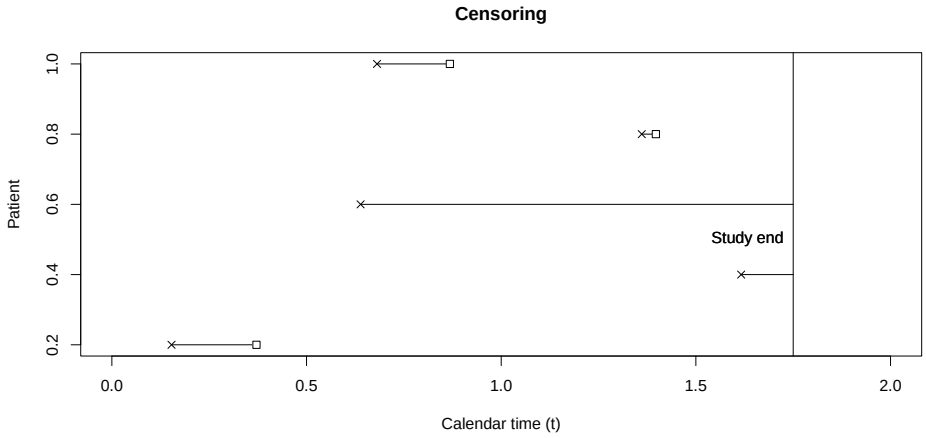

Figure 3.2: Example of generalized Type I censoring, where each individual has a separate study entry time.

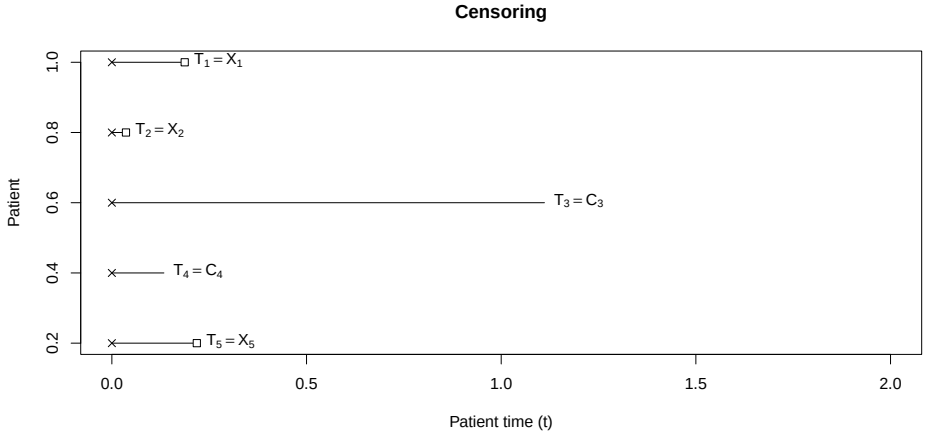

Figure 3.3: Example of generalized Type I censoring, viewed in patient time.

- $\delta_i = 1 \implies T_i = X_i$.

This is different from Type I censoring in that each individual has a different censoring time.

3.1.3 Type II censoring

Type II censoring occurs when all units have the same study entry time, but researchers design the study to end when $r < n$ units fail out of n total units under observation.

- For the first r, lucky or unlucky participants, $\delta_i = 1 \implies T_i = X_{(i)}$ or the i^{th} order statistic.
- For the remaining $n - r$ individuals, $\delta_i = 0 \implies X_i > X_{(r)}$.

3.1.4 Generalized Type II censoring

You may be wondering, what happens when units have differing start times but we want to end the trial after the r-th failure? It turns out that this was not a solved problem until Rühl et al. 2023, which was quite surprising to me.

3.1.5 Independent censoring

A third type of censoring, helpfully called independent censoring, takes $X_i \perp\!\!\!\perp C_i$, and thus conclusions similar to those of generalized type I censoring can be drawn:

- For all individuals such that $\delta_i = 0 \implies X_i > C_i$
- $\delta_i = 1 \implies T_i = X_i$.

3.2 Noninformative censoring

All of the previous censoring scenarios can be summarized as noninformative censoring. Let the parameters indexing the censoring distribution be ϕ, while the parameters indexing the failure time distribution are θ. Noninformative censoring is defined as the following equality:

$$\lambda_{X_i}(t) = \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathbb{P}_{\theta,\phi} \left(t \le X_i < t + \Delta t \mid X_i \ge t, C_i \ge t \right) \tag{3.1}$$

Note that this implies the following:

$$\mathbb{P}_{\theta} \left(t \le X_i < t + \Delta t \mid X_i \ge t \right) = \mathbb{P}_{\theta,\phi} \left(t \le X_i < t + \Delta t \mid X_i \ge t, C_i \ge t \right). \tag{3.2}$$

which is equivalent to writing that failure

For independent censoring, Equation (3.2) holds, given that $P_{\theta,\phi}(X_i > t, C_i > c) = P_\theta(X_i > t)P_\phi(C_i > c)$ and that θ and ϕ are *variationally independent*.

This means that the parameter space $\Omega_{\theta,\phi}$ is the Cartesian product of the parameter space for θ and ϕ.

Definition 3.2.1. Variational independence Let $\theta \in \Omega_\theta$ and $\phi \in \Omega_\phi$. The joint space is denoted as $\Omega_{\theta,\phi}$. If $\Omega_{\theta,\phi} = \Omega_\theta \times \Omega_\phi$, θ and ϕ are variationally independent. In other words, the range for θ does not change given a value for ϕ.

Under independent censoring, the observable hazard for uncensored failure times is as follows:

$$\frac{-\frac{\partial}{\partial u}P_{\theta,\phi}(X_i > u, C_i > t)\mid_{u=t}}{P_{\theta,\phi}(X_i > t-, C_i > t-)} = \frac{-\frac{d}{du}S_{X_i}(u;\theta)}{S_{X_i}(t-;\theta)} \tag{3.3}$$

Here's an example that demonstrates the nonidentifiability of the joint distribution for censoring and failure time:

Example 3.2.1. Dependent failure and censoring
Let $P_{\theta,\alpha,\mu}(X_i > x, C_i > c) = \exp(-\alpha x - \mu c - \theta x c)$. We can find the marginal survival functions just by evaluating $P_{\theta,\alpha,\mu}(X_i > x, C_i > 0)$ and vice-versa, which yields:

$$P_\alpha(X_i > x) = \exp(-\alpha x)$$
$$P_\mu(C_i > c) = \exp(-\mu c)$$

Both of these distributions have constant hazards. However, the observable hazard is the following:

$$\frac{-\frac{\partial}{\partial u}P_{\theta,\alpha,\mu}(X_i > u, C_i > t)\mid_{u=t}}{P_{\theta,\alpha,\mu}(X_i > t-, C_i > t-)} = \alpha + \theta t$$
$$\frac{-\frac{\partial}{\partial u}P_{\theta,\alpha,\mu}(X_i > t, C_i > u)\mid_{u=t}}{P_{\theta,\alpha,\mu}(X_i > t-, C_i > t-)} = \mu + \theta t$$

This leads to an observable survival function:

$$S_{X_i}(x;\alpha,\theta) = \exp(-\alpha x - \theta x^2/2)$$
$$S_{C_i}(c;\mu,\theta) = \exp(-\mu c - \theta c^2/2)$$

If we mistakenly assume that the failure time and the censoring time are independent we'll get the following joint distribution:

$$S_{X_i}(x;\alpha,\theta)S_{C_i}(c;\mu,\theta) \neq \exp(-\alpha x - \mu c - \theta x c).$$

However, if we calculate the true observable survival function $P_{\theta,\alpha,\mu}(X_i > x, C_i > X_i)$ we get:

$$\int_x^\infty -\frac{\partial}{\partial u} P_{\theta,\alpha,\mu}(X_i > u, C_i > t) \mid_{u=t} dt = \int_x^\infty (\alpha + \theta t) \exp(-\alpha t - \mu t - \theta t^2) dt$$

while the observable survival function implied the by the erroneously assumed independent distributons is:

$$\begin{aligned}\int_x^\infty -\frac{\partial}{\partial u} S_{X_i}(x;\, \alpha, \theta) S_{C_i}(c;\, \mu, \theta) \mid_{u=t} dt &= \int_x^\infty (-\frac{d}{dt} \exp(-\alpha t - \theta t^2/2)) \exp(-\mu t - \theta t^2/2) \\ &= \int_x^\infty (\alpha + \theta t) \exp(-\alpha t - \mu t - \theta t^2) du\end{aligned}$$

Thus, two different joint densities lead to the same observable survival functions, so the joint distribution is nonidentifiable.

Here is an example showing that we may have dependent censoring and failure times, but still end up with noninformative censoring:

Example 3.2.2. Dependent failure and censoring can be noninformative
Let Y_1, Y_2 and Y_{12} be exponentially distributed with rates $\alpha_1, \alpha_2, \alpha_{12}$, respectively. Let $X = Y_1 \wedge Y_{12}$ and $C = Y_2 \wedge Y_{12}$. The survival function $P_{\alpha_1,\alpha_2,\alpha_{12}}(X > x, C > c) = P(Y_1 > x, Y_2 > c, Y_{12} > x \vee c) = e^{-\alpha_1 x - \alpha_2 c - \alpha_{12} x \vee c}$ Then marginally X is exponential with rate $\alpha_1 + \alpha_{12}$, which is also equal to its hazard function. In order for noninformative censoring to hold, we need to check Equation (3.1), or that

$$\alpha_1 + \alpha_{12} = \lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathbb{P}_{\alpha_1,\alpha_2,\alpha_{12}} \left(t \le X < t + \Delta t \mid X \ge t, C \ge t\right) \tag{3.4}$$

Because $t + \Delta t \vee t = t + \Delta t$ as $\Delta t > 0$,

$$\lim_{\Delta t \searrow 0} \frac{e^{-\alpha_1 t - \alpha_2 t - \alpha_{12} t} - e^{-(\alpha_1 + \alpha_{12})(t + \Delta t) - \alpha_2 t}}{\Delta t} \tag{3.5}$$

which just equals $e^{-\alpha t} - \frac{d}{ds} e^{-(\alpha_1 + \alpha_{12})s} \mid_{s=t}$ or $(\alpha_1 + \alpha_{12}) e^{-\alpha_1 t - \alpha_2 t - \alpha_{12} t}$. Then

$$\lim_{\Delta t \searrow 0} \frac{1}{\Delta t} \mathbb{P}_{\alpha_1,\alpha_2,\alpha_{12}} \left(t \le X < t + \Delta t \mid X \ge t, C \ge t\right) = \frac{(\alpha_1 + \alpha_{12}) e^{-\alpha_1 t - \alpha_2 t - \alpha_{12} t}}{e^{-\alpha_1 t - \alpha_2 t - \alpha_{12} t}} \tag{3.6}$$

$$= \alpha_1 + \alpha_{12} \tag{3.7}$$

So in this case, while X and C are dependent, we still have noninformative censoring.

The benefit of noninformative censoring is that we can ignore the censoring random variables when constructing the likelihood for the survival random variables.

3.2.1 Reasons for informative censoring

A simple hypothetical situation with informative censoring would be one in which sick patients are lost to follow-up.

3.3 Truncation

While censoring can be seen as partial information about an observation, truncation deals with exact observations of selected units. The simplest example of truncation is when measurements are made using an instrument with a lower limit of detection. Imagine using a microscope to measure the diameter of cells on a plate that has a lower limit of detection of 5 microns. If interest lies in inferring the population mean diameter of the cells, one must take into account the fact that only cells with diameters of greater than 5 micros can be seen with the microscope.

Failure to take truncation into account can be a source of bias in inference.

$$\begin{aligned}\mathbb{E}[X_i] &= \mathbb{E}[X_i \mid X_i \geq V]\, P(X_i \geq V) + \mathbb{E}[X_i \mid X_i < V]\, P(X_i < V) \\ &= \mathbb{E}[X_i \mid X_i \geq V] + P(X_i < V)(\mathbb{E}[X_i \mid X_i < V] - \mathbb{E}[X_i \mid X_i \geq V]) \\ &\leq \mathbb{E}[X_i \mid X_i \geq V]\end{aligned}$$

The last line follows because $(\mathbb{E}[X_i \mid X_i < V] - \mathbb{E}[X_i \mid X_i \geq V]) \leq 0$. Using an estimator for $\mathbb{E}[X_i \mid X_i \geq V]$ when the target of inference in $\mathbb{E}[X_i]$ would result in positive bias. Of course, when the estimator instead estimates $\mathbb{E}[X_i \mid X_i < V]$ the bias would be negative. Depending on the value of V and the distribution of X_i, the bias can be severe.

For example, suppose a researcher is interested in learning about the impact of medication refills on the lifespans of patients. The researcher has access to a database in which they select patients who refilled medications at least once. The researcher subsequently selects a control group that is perfectly matched to the medication refill group, and upon analyzing the data, the analyst discovers that refilling prescription medication leads to longer lifespans. What is wrong with this analysis?

The observations in this example can be said to be left-truncated, because the researcher conditions the observations in the treatment group on having a lifespan long enough to fill a medication.

Formally, we say that the density for a truncated observation is conditioned on the probability of the observation lying in the truncated region.

- If a researcher selects $\mathbb{1}(X_i \geq V)$ we say the data are left-truncated, and $f_{X_i}(x;\theta) = \frac{-\frac{d}{dx} S_{X_i}(x;\theta)}{S_{X_i}(v;\theta)}$

- If a researcher selects $\mathbb{1}(X_i \leq U)$ we say the data are right-truncated, and $f_{X_i}(x;\theta) = \frac{-\frac{d}{dx}S_{X_i}(x;\theta)}{1-S_{X_i}(u;\theta)}$
- If a researcher selects $\mathbb{1}(V \leq X_i \leq U)$ we say the data are interval-truncated, and $f_{X_i}(x;\theta) = \frac{-\frac{d}{dx}S_{X_i}(x;\theta)}{S_{X_i}(v;\theta)-S_{X_i}(u;\theta)}$

Bibliography

[1] John P Klein, Melvin L Moeschberger, et al. *Survival analysis: techniques for censored and truncated data*. Vol. 1230. Springer, 2003.

[2] Odd Aalen, Ornulf Borgan, and Hakon Gjessing. *Survival and event history analysis: a process point of view*. Springer Science & Business Media, 2008.

[3] Thomas R Fleming and David P Harrington. "Counting Processes and Survival Analysis". In: *Wiley Series in Probability and Statistics* (2005).

[4] Sidney Resnick. *A probability path*. Springer, 2019.

[5] Jasmin Rühl, Jan Beyersmann, and Sarah Friedrich. "General independent censoring in event-driven trials with staggered entry". en. In: *Biometrics* 79.3 (2023), pp. 1737–1748. ISSN: 0006-341X, 1541-0420. DOI: 10.1111/biom.13710.