2 Basic Concepts of Failure Time Analysis

2.1 Continuous Time

Time is represented by the nonnegative random variable T with cumulative density function

$$F(t) = P(T \leq t),$$

and density

$$f(t) = dF(t)/dt.$$

For failure time analysis, T is generally characterized by other quantities. The survivor function gives the probability that T exceeds t:

$$G(t) = P(T > t) = 1 - F(t).$$

It always holds that no individual has failed at $T = 0$

$$G(0) = 1, \tag{2.1}$$

and it is usually assumed that every subject will fail eventually

$$\lim_{t \to \infty} G(t) = 0. \tag{2.2}$$

Variables with survivor function not satisfying 2.2 are called defective, for those it follows that E[T]=∞. The probability of failure in the small interval [t,t+dt) can be approximated by $h(t)dt$ (Aalen et al. 2008, pp. 5–17). The function $h(t)$ is the hazard rate, defined as:

$$h(t) = \lim_{\Delta \to 0} \frac{P(t \leq T < t + \Delta | T \geq t)}{\Delta}. \tag{2.3}$$

M. Kaeding, *Bayesian Analysis of Failure Time Data Using P-Splines,* BestMasters,
DOI 10.1007/978-3-658-08393-9_2,

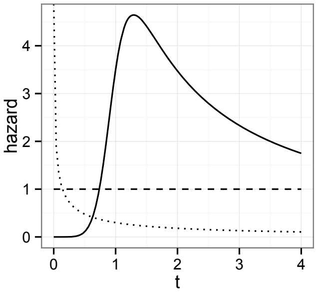

Figure 2.1: Some hazard rates

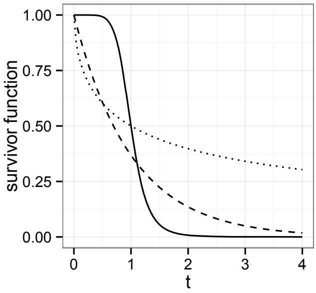

Figure 2.2: Some survivor functions

The probability $P(t \leq T < t+\Delta | T \geq t)$ is

$$\frac{F(t+\Delta)-F(t)}{G(t)}.$$

Hence 2.3 is

$$\frac{1}{G(t)} \lim_{\Delta \to 0} \frac{F(t+\Delta)-F(t)}{\Delta} = \frac{F'(t)}{G(t)} = \frac{f(t)}{G(t)},$$

showing that the hazard is a conditional density.

The cumulative hazard rate is

$$H(t) = \int_0^t h(t) = \int_0^t \frac{f(u)}{G(u)}\, du = [-\log G(u)]_0^t = -\log G(t),$$

due to 2.1. Hence, the survivor function can be written in terms of the hazard rate:

$$G(t) = \exp(-\int_0^t h(t)) = \exp(-H(t)). \tag{2.4}$$

The same applies for the density:

$$f(t) = h(t)G(t) = h(t)\exp(-H(t)).$$

Because of these relationships, the random variable T is fully specified by one of the given quantities. From 2.4, it can be seen that the function $h(t)$ only needs to satisfy

$$\int_0^t h(s)\,ds < \infty,$$

for all t and

$$\int_0^\infty h(s)\,ds = \infty$$

to be the hazard rate of a nondefective continuous variable (Kalbfleisch and Prentice 2002, p. 9). Many models in failure time modeling are formulated in terms of the hazard rate first.

2.2 Discrete Time

In the case of grouped failure times, an unobservable continous random variable $T^\star$ is partitioned into m+1 intervals $[a_0 = 0, a_1), [a_1, a_2), ..., [a_m, a_{m+1} = \infty)$, (Lawless 2003, p. 370). Observed are discrete failure times from the random variable $T = \{1, 2, .., m+1\}$, so that T$= t$ corresponds to T$^\star \in [a_{t-1}, a_t)$. The hazard in terms of T is

$$h(t) = P(T = t|T \geq t) = \frac{P(T = t)}{P(T \geq t)} = \frac{f(t)}{G(t-1)}. \tag{2.5}$$

Expressing 2.5 in terms of $T^\star$ gives:

$$h(t) = \frac{G^\star(a_t) - G^\star(a_{t-1})}{G^\star(a_t)} = 1 - \exp(-\int_{a_{t-1}}^{a_t} h^\star(u)\,du).$$

This is the probability of failure in interval t, conditional on reaching the interval. A discrete time model can be specified in terms of T or T*. Failure after interval t is a result of a sequence of binary trials unfolding in time (Kalbfleisch and Prentice 2002, p. 9):

$$G(t) = P(T > t) = P(T \neq 1 \cap T \neq 2 ... \cap T \neq t) =$$

$$P(T \neq 1)P(T \neq 2|T \neq 1)P(T \neq 3|T \neq 1, T \neq 2)...P(T \neq t|T \neq 1, ..., T \neq t-1).$$

The probability $P(T \neq x|T \neq x-1)$ is given by $1 - h(x)$, it follows that in analogy to the continuous case the survivor function can be expressed in terms of the hazard rate:

$$G(t) = \prod_{j=1}^{t} (1 - h(j)).$$

Assuming grouped failure times might not be appropriate in all cases, as some random variables are intrinsically discrete. Some helpful results follow from this assumption however, and estimation is easier by deriving inferences on the likelihood contributions following from 2.2, leading to an identical modeling framework.

2.3 Likelihood Construction

Failure time data have some special characteristics which have to be accounted for in the construction of the likelihood. A failure time is referred to as censored when the actual failure time is not observed but it is only known to fall into an interval. Failure times are *left-truncated* if they are only observable if they exceed a truncation time. Time varying covariates are often available in the data set. In the following sections, based on Klein and Moeschberger (2003, pp. 63-77), it will be clarified how these conditions are accounted for in the formulation of the likelihood. Conceptually, these adjustments can be represented in an unified framework by varying the likelihood contributions. As a consequence, the likelihood becomes more difficult to work with but there are computational methods which simplify estimation.

2.3.1 Censoring and Truncation

In the presence of right-censoring, the observed failure time for an individual is

$$t_i = \min(\dot{t}_i, c_i).$$

Here, $\dot{t}_i$ is the true failure time and c_i is the censoring time. The indicator variable v_i is defined as

$$v_i = \begin{cases} 1 & \text{if } t_i \leq c_i, \\ 0 & \text{if } t_i > c_i, \end{cases}$$

and is usually referred to as censoring indicator. The available data is given by:

$$D = \{(t_i, \delta_i, \boldsymbol{z}_i^\top)_{i=1}^n\}.$$

Here, $\boldsymbol{z}_i = (z_{i1}, z_{i2}, ..)^\top$ is the vector of covariates of individual i. To proceed, it is necessary to make assumptions about the process generating the censoring times. Under the assumption of *random censoring*, $C_1, ..., C_n$ are i.i.d. random variables with survivor function S() and pdf s() depending on $\boldsymbol{\phi}$, independent of $T_1, ..., T_n$ and each other. Let $\boldsymbol{\theta}$ be the parameter vector of interest on which the survivor function of $T_1, ..., T_n$ depends. Under random censoring, the full likelihood contribution $L_i^\star$ is

$$L_i^\star = G(t_i|\boldsymbol{z}_i) = P(T_i > t_i|\boldsymbol{z}_i)s(t_i)$$

for a right-censored failure time and

$$L_i^\star = G(t_i|\boldsymbol{z}_i)h(t_i|\boldsymbol{z}_i)S(t_i) = f(t_i|\boldsymbol{z}_i)S(t_i)$$

for a completely observed failure time. Under *noninformative censoring*, we have $G(t_i|\boldsymbol{\theta}, \boldsymbol{\phi}) = G(t_i|\boldsymbol{\theta})$. Further we assume that $T_1, ..., T_n$ are i.i.d. or independent given the covariates. Under those assumptions, the likelihood is given by:

$$\begin{aligned} L(\boldsymbol{\theta}|D) &= c \prod_{i=1}^{n} h(t_i|\boldsymbol{z}_i, \boldsymbol{\theta})^{v_i} G(t_i|\boldsymbol{z}_i, \boldsymbol{\theta}) \\ &\propto \prod_{i=1}^{n} L_i, \end{aligned}$$

where $c = \prod_{i=1}^n S(t_i)^{v_i} s(t_i)^{1-v_i}$ is a multiplicative constant and

$$L_i = h(t_i|\boldsymbol{z}_i, \boldsymbol{\theta})^{v_i} G(t_i|\boldsymbol{z}_i, \boldsymbol{\theta}).$$

For Bayesian analysis, the assumption $f(\boldsymbol{\theta}, \boldsymbol{\phi}) = f(\boldsymbol{\theta}) f(\boldsymbol{\phi})$ is also necessary so that $f(\boldsymbol{\phi})$ factors out of the posterior. In this thesis it is always assumed that censoring is random and noninformative. For discrete failure time data a failure indicator y is introduced for every interval before and including the failure time, so that $P(y_{ij} = h(j|\mathbf{z}_i))$:

$$y_{ij} = \begin{cases} 1 & \text{if individual i fails in interval } [a_{j-1}, a_j), \\ 0 & \text{if } t_i > a_j. \end{cases}$$

Assuming that censoring occurs at the end of the interval, the likelihood contribution of an uncensored and a right-censored individual respectively equal

$$\prod_{i=1}^{t_i} P(y_{ij} = 0)^{1-y_{ij}} P(y_{ij} = 1)^{y_{ij}} = h(t_i|\mathbf{z}_i) \prod_{i=1}^{t_{i-1}} (1 - h(i|\mathbf{z}_i)) =$$
$$h(t_i|\mathbf{z}_i) G(t_i - 1|\mathbf{z}_i) = f(t_i|\mathbf{z}_i),$$

and

$$\prod_{i=1}^{t_i} P(y_{ij} = 0) = \prod_{i=1}^{t_i} (1 - h(j|\mathbf{z}_i)) = G(t_i|\mathbf{z}_i).$$

The likelihood is:

$$L(\theta|D) = \prod_{i=1}^{n} \prod_{j=1}^{t_i} (1 - h(j|\mathbf{z}_i))^{1-y_{ij}} h(j|\mathbf{z}_i)^{y_{ij}},$$

which is the likelihood of a Bernoulli distribution. The same result would have been obtained for an intrinsically discrete random variable. Discrete failure times can be analyzed by methods for this distribution. In practice, this is achieved by changing a data set given in the usual form for failure time analysis 2.1 into the longitudinal form 2.2. Under interval censoring, it is only known that failure occurred

Table 2.1: Discrete data

id	t	z1	v
1	3	9	1
2	2	12	0

during the interval $[l_i, r_i)$. Interval-censoring can be viewed as generalization of right-censoring, the interval corresponding to right-censoring is $[l_i, \infty)$, while by

Table 2.2: Discrete data - longitudinal

id	y	z1
1	0	9
1	0	9
1	1	9
2	0	12
2	0	12

convention the interval of a uncensored individual is $[l_i = t_i, r_i = t_i)$. For continuous time, the interval can be set to $(l_i, r_i]$ $[l_i, r_i]$ or (l_i, r_i) as the same information about the failure time is represented (Sun 2006, p. 15), this is not true for discrete time. In this thesis, the notation $[l_i, r_i)$ is used. Grouped failure times are a special case of interval-censoring where all no intervals overlap (Lawless 2003, p. 64). The available data is given by

$$D = \{([l_i, r_i), \boldsymbol{z}_i^\top)_{i=1}^n\}.$$

The likelihood contribution is given by:

$$L_i = P(l_i \leq T_i < r_i|\boldsymbol{z}_i) = G(l_i|\boldsymbol{z}_i) - G(r_i|\boldsymbol{z}_i).$$

Under left-truncation, a failure time can only be observed if it exceeds a truncation time tr_i. Analysis of those cases proceeds by conditioning on failure after tr_i. For example, for continuous time, the contribution of an uncensored, left-truncated individual is:

$$\begin{aligned}\frac{f(t_i|\boldsymbol{z}_i)}{S(tr_i|\boldsymbol{z}_i)} &= \frac{h(t|\boldsymbol{z}_i)\exp(-\int_0^{t_i} h(t|\boldsymbol{z}_i)\,dt}{\exp(-\int_0^{tr_i} h(t|\boldsymbol{z}_i)\,dt} \\ &= h(t|\boldsymbol{z}_i)\exp(-\int_{tr_i}^{t_i} h(t|\boldsymbol{z}_i)\,dt.\end{aligned}$$

For discrete time:

$$\frac{f(t_i|z_i)}{S(tr_i|z_i)} = \frac{h(t_i|z_i)\prod_{j=1}^{t_i-1}(1-h(j|z_i))}{\prod_{j=1}^{tr_i}(1-h(j|z_i))} = h(t_i|z_i)\prod_{j=tr_i+1}^{t_i-1}(1-h(j|z_i)),$$

so conveniently by deleting failure indicators up to and including the truncation time, the likelihood contribution is correct. Combining the concepts, the data is given by

$$\{(l_i, r_i, tr_i, z_i)_{i=1}^n\}.$$

2.3.2 Time Varying Covariates

For covariates depending on time a distinction must be made between *internal* and *external* covariates (Kalbfleisch and Prentice 2002, pp. 196–199): For the former case it holds that

$$h(t|\boldsymbol{Z}(t), \boldsymbol{\beta}, T \geq u) = h(t|\boldsymbol{Z}(t), \boldsymbol{\beta}, T = u), 0 < u \leq t,$$

this implies that the covariates $\boldsymbol{Z}(t)$ affect the hazard, but failure does not affect the covariate path $\boldsymbol{Z}(t)$. Internal covariates are those that are directly involved with failure: As such, $G(t|\boldsymbol{Z}(t), \boldsymbol{\beta})$ can no longer be interpreted as a survivor function. Inclusion of internal covariates is problematic and here attention is restricted to external covariates. For example in the context of unemployment durations, an example of an internal covariate would be the amount of unemployment benefits an individual receives. Given that the amount is > 0 at month t, we have[1] $G(T > t|\boldsymbol{Z}(t)) = 1$. A (strictly seen, approximate) external variable might be the current rate of unemployment. Treatment of time varying covariates proceeds by taking them to be a stochastic process, an important property of which is *predictability*; informally, covariates are predictable if the values which explain variation in the hazard rate at time t are (to the researcher) known at an infinitesimal short moment before t (Berg 2001). The author gives as an example the case of an individual making a decision, e.g. accepting a job offer, under the anticipation of

[1] A model could in fact not even be fit with usual procedures here because of perfect seperation.

the realization of T. If this is unknown to the analyst, predictability is not given. In this thesis it is always assumed that covariates are predictable. Under this assumption - given regularity conditions - standard methods can be used with time varying covariates for relative risk models introduced in the next section.

While general sampling paths are possible for $\boldsymbol{Z}(t)$, changes in covariates values are usually observed at discrete points $(t_0 = 0 < t_1 < \ldots < t_{n_{i,k}})$. The survivor function of an individual can be written as product of conditional survivor functions without changing the likelihood:

$$
\begin{aligned}
G(t_i|\boldsymbol{z}_i(t)) = \exp\{-\int_0^t h(u|\boldsymbol{z}_i(u))\,du\} = \exp\{-\int_{t_0}^{t_1} h(u|\boldsymbol{z}_i(u))\,du - \\
\int_{t_1}^{t_2} h(u|\boldsymbol{z}_i(u))\,du - \ldots - \int_{t_{n_k}-1}^{t_{n_{i,k}}} h(u|\boldsymbol{z}_i(u))\,du\} \\
= \prod_{j=1}^{n_{i,k}} G(t_j|t_j > t_{j-1}, \boldsymbol{z}_i(j)).
\end{aligned}
$$

Every term $G(t_j|t_j > t_{j-1}, \boldsymbol{z}_i(j))$ is the likelihood contribution of an right-censored individual with covariate $\boldsymbol{z}_i(j)$ and failure time t_j left-truncated at t_{j-1}. As a consequence, the data set can be adjusted by a process called *episode splitting* (Blossfeld and Rohwer 2002, pp. 140–142), which can be seen in table 2.3. On the left is the

Table 2.3: Episode splitting

id	t	v	z1	tl	id	t	v	z1	z2	tl
1	5	1	3	0	1	3	0	3	11	0
2	7	0	5	0	1	5	1	3	15	3
					2	2	0	5	9	0
					2	5	0	5	4	2
					2	7	0	5	6	5

data set before, on the right the data set after episode splitting. After the split, the time varying covariate z_2 can be included. For discrete time, time-varying covariates can be inluded by varying the covariate values across the intervals.

2.4 Relative Risk and Log-Location-Scale Family

The Cox model or relative risk model, due to Cox (1972) is probably the most often used model for failure time modeling in continuous time. For models belonging to the relative risk family, the hazard rate is specified as:

$$h(t|\boldsymbol{z}(t),\boldsymbol{\beta}) = h_0(t)\exp(\boldsymbol{\beta}^T\boldsymbol{z}(t)), \tag{2.6}$$

implying a loglinear model for the hazard rate:

$$\log h(t|\boldsymbol{z}(t),\boldsymbol{\beta}) = \log h_0(t) + \boldsymbol{z}(t)^\top\boldsymbol{\beta}.$$

Here, the hazard rate consists of the baseline hazard $h_0(t)$, corresponding to $\boldsymbol{z} = \mathbf{0}$ on which the covariates act multiplicatively by $\exp(\boldsymbol{\beta}^\top\boldsymbol{z}(t))$. For time-constant covariates the model 2.6 is also known as proportional hazards model. In this case, the ratio of hazards of two individuals with covariate vectors $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$ is

$$\frac{h_i(t|\boldsymbol{z}_i,\beta)}{h_j(t|\boldsymbol{z}_j,\beta)} = \frac{h_0(t)\exp(\boldsymbol{\beta}^\top\boldsymbol{z}_i)}{h_0(t)\exp(\boldsymbol{\beta}^\top\boldsymbol{z}_j)} = \exp(\boldsymbol{\beta}^\top(\boldsymbol{z}_i-\boldsymbol{z}_j)), \tag{2.7}$$

so $h_i(t|\boldsymbol{z}_i) \propto h_j(t|\boldsymbol{z}_j)$, independent of t. A one unit change in covariate z_p - given the other covariates are fixed - corresponds to multiplication of the hazard rate by the factor $\exp(\beta_p)$. This factor is also called the *hazard ratio* or *relative risk* of covariate p (Aalen et al. 2008, p. 9) as it gives the ratio 2.7 if $\boldsymbol{z}_i - \boldsymbol{z}_j = (0,..,0,1,0,..,0)^\top$, where the one is in position p. Frequentist inference is often based on the partial likelihood, which is free of the baseline hazard:

$$L_p(\beta|\mathscr{D}) = \prod_{i=1}^{d}\frac{\exp(\boldsymbol{z}_i\boldsymbol{\beta})}{\sum_{j\in\mathscr{R}(i)}\exp(\boldsymbol{z}_j\boldsymbol{\beta})}, \tag{2.8}$$

$t_{(1)},t_{(2)},\dots,t_{(d)}$ are the ordered failure times and

$$\mathscr{R}(x) = \{i : t_i \geq x\}$$

http://www.springer.com/978-3-658-08392-2

Bayesian Analysis of Failure Time Data Using P-Splines
Kaeding, M.
2015, IX, 110 p. 23 illus., Softcover
ISBN: 978-3-658-08392-2