BIO 244: Unit 1

Survival Distributions, Hazard Functions, Cumulative Hazards

1.1 Definitions:

The goals of this unit are to introduce notation, discuss ways of probabilistically describing the distribution of a 'survival time' random variable, apply these to several common parametric families, and discuss how observations of survival times can be right-censored.

Suppose T is a non-negative random variable representing the time until some event of interest. For example, T might denote:

- the time from diagnosis of a disease until death,
- the time between administration of a vaccine and development of an infection,
- the time from the start of treatment of a symptomatic disease and the suppression of symptoms.

We shall assume that T is continuous unless we specify otherwise. The probability density function (pdf) and cumulative distribution function (cdf) are most commonly used to characterize the distribution of any random variable, and we shall denote these by $f(\cdot)$ and $F(\cdot)$, respectively:

$$\left.\begin{array}{lll} pdf: & f(t) & \\ cdf: & F(t) & = P(T \leq t) \end{array}\right\} \quad F(0) = P(T = 0).$$

Because T is non-negative and usually denotes the elapsed time until an event, it is commonly characterized in other ways as well:

Survivor function:

$$S(t) \stackrel{\text{def}}{=} 1 - F(t) = P(T > t) \qquad \text{for} \quad t \geq 0.$$

The survivor function simply indicates the probability that the event of interest has not yet occurred by time t; thus, if T denotes time until death, $S(t)$ denotes probability of surviving beyond time t.

Note that, for an arbitrary T, $F(\cdot)$ and $S(\cdot)$ as defined above are right continuous in t. For continuous survival time T, both functions are continuous in t. However, even when $F(\cdot)$ and $S(\cdot)$ are continuous, the nonparametric estimators, say $\hat{F}(\cdot)$ and $\hat{S}(\cdot)$, of these that we will consider are discrete distributions. For example, $\hat{F}(\cdot)$ might be the c.d.f. corresponding to the discrete distribution that places mass $m_1, m_2, \cdots, m_k$ at certain times $\tau_1, \tau_2, \cdots, \tau_k$. Thus, even though $F(\cdot)$ is continuous, its estimator $\hat{F}(\cdot)$ is (only) right continuous, and thus its value at a certain time point, say τ_2, will be $m_1 + m_2$ if we define the c.d.f. to be right continuous (but equal to m_1 if we had defined the c.d.f. to be left continuous).

Hazard function:

$$\begin{aligned} h(t) &\stackrel{\text{def}}{=} \lim_{h \downarrow 0} \frac{P[t \leq T < t + h | T \geq t]}{h} \\ &= \frac{f(t)}{S(t-)} \end{aligned}$$

with $S(t-) = \lim_{s \uparrow t} S(s)$. That is, the hazard function is a conditional density, given that the event in question has not yet occurred prior to time t. Note that for continuous T, $h(t) = -\frac{d}{dt} \ln[1 - F(t)] = -\frac{d}{dt} \ln S(t)$.

Cumulative hazard function:

$$H(t) \stackrel{\text{def}}{=} \int_0^t h(u) du \qquad t \geq 0.$$

For continuous T,

$$H(t) = -\ln[1 - F(t)] \;=\; -\ln S(t).$$

Note that for continous T,

$$\begin{aligned} S(t) &= e^{-H(t)} \\ f(t) &= h(t)e^{-H(t)}. \end{aligned}$$

Note 1: Note that h(t)dt = f(t)dt/S(t) $\approx$ P[fail in [t,t+dt) | survive until t]. Thus, the hazard function might be of more intrinsic interest than the p.d.f. to a patient who had survived a certain time period and wanted to know something about their prognosis.

Note 2: There are several reasons why it is useful to introduce the quantities h(t) and H(t):

- Interpretability: Suppose T denotes time from surgery for breast cancer until recurrence. Then when a patient who had received surgery visits her physician, she would be more interested in conditional probabilities such as "Given that I haven't had a recurrence yet, what are my chances of having one in the next year" than in unconditional probabilities (as described by the p.d.f.).
- Analytic Simplifications: When the data are subject to right censoring, hazard function representations often lead to easier analyses. For example, imagine assembling a cohort of N men who just have turned 50 years of age and then following them for 1 year. Then if d of the men die during the year of follow-up, the ratio d/N estimates the (discrete) hazard function of T=age at death. We will see that $H(\cdot)$ has nice analytical properties.
- Modeling Simplifications: For many biomedical phenomena, T is such that h(t) varies rather slowly in t. Thus, $h(\cdot)$ is well-suited for modeling.

Note 3: It is useful to think about real phenomena and how their hazard functions might be shaped. For example, if T denotes the age of a car when it

first has a serious engine problem, then one might expect the corresponding hazard function h(t) to be increasing in t; that is, the conditional probability of a serious engine problem in the next month, given no problem so far, will increase with the life of the car. In contrast, if one were studying infant mortality in a region of the world where there was poor nutrition, one might expect h(t) to be decreasing during the first year of life. This is known to be due to selection during the first year of life. Finally, in some applications (such as when T is the lifetime of a light bulb), the hazard function will be approximately constant in t. This means that the chances of failure in the next short time interval, given that failure hasn't yet occurred, does not change with t; e.g., a 1-month old bulb has the same probability of burning out in the next week as does a 5-year old bulb. As we will see below, this 'lack of aging' or 'memoryless' property uniquely defines the exponential distribution, which plays a central role in survival analysis.

1.2 Common Families of Survival Distributions

Exponential Distribution: denoted $T \sim Exp(\lambda)$. For $t > 0$,

$$\begin{aligned} f(t) &= \lambda e^{-\lambda t} \quad \text{for } \lambda > 0 \text{ (scale parameter)} \\ F(t) &= 1 - e^{-\lambda t} \quad S(t) = e^{-\lambda t} \\ h(t) &= \lambda \quad \leftarrow \boxed{\text{constant hazard function}} \\ H(t) &= \lambda t \end{aligned}$$

characteristic function:

$$\phi(u) = E[e^{iuT}] = \frac{\lambda}{\lambda - iu}$$

$$\therefore \quad E[T^r] = \left. \frac{\phi^{(r)}(u)}{i^r} \right|_{u=0}$$

- $E(T) \;=\; \dfrac{1}{\lambda}$

- $V(T) \;=\; \dfrac{1}{\lambda^2}$

- "Lack of Memory": $P[T > t] \;=\; P[T > t + t_0 | T > t_0]$ for any $t_0 > 0$ (probability of surviving <u>another</u> t time units does not depend on how long you've lived so far)

- Also, the exponential family is closed to scale changes; that is: $T \sim Exp(\lambda),\ c > 0 \Rightarrow c \cdot T \sim Exp(\lambda/c)$.

2-Parameter Gamma Distribution:

The 2-parameter gamma distribution, which is denoted $G(\alpha, \lambda)$, can be viewed as a generalization of the exponential distribution. It arises naturally (that is, there are real-life phenomena for which an associated survival distribution is approximately Gamma) as well as analytically (that is, simple functions of random variables have a gamma distribution).

$$f(t) \;=\; \frac{\lambda^\alpha t^{\alpha-1} e^{-\lambda t}}{\Gamma(\alpha)} \;\text{ for }\; t > 0.$$

Parameters $\lambda > 0$ and $\alpha > 0$

$$\Gamma(\alpha) \;=\; \text{gamma func.} \;=\; \int_0^\infty t^{\alpha-1} e^{-t}\, dt.$$

- characteristic function: $\phi(u) \;=\; \left(\dfrac{\lambda}{\lambda - iu}\right)^\alpha$

- $E(T) \;=\; \dfrac{\alpha}{\lambda}$

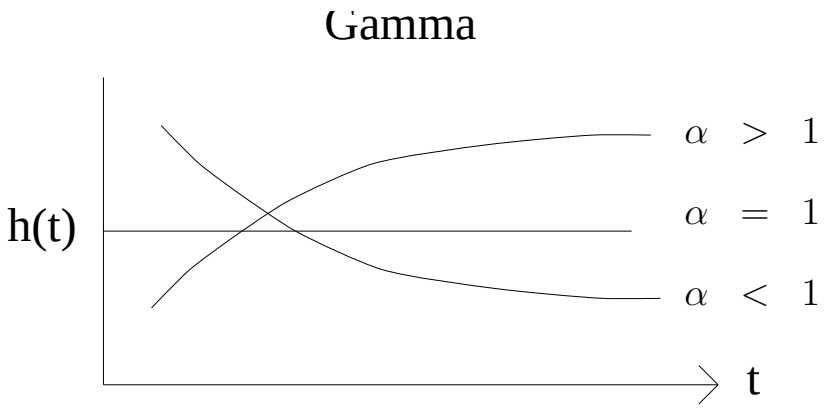
- $V(T) = \dfrac{\alpha}{\lambda^2}$

- $G(1, \lambda) = Exp(\lambda)$

- $T_1 \sim G(\alpha_1, \lambda),\ \ T_2 \sim G(\alpha_2, \lambda),\ \ T_1 \perp T_2 \Longrightarrow T_1 + T_2 \sim G(\alpha_1 + \alpha_2, \lambda)$

- if $\alpha = \frac{k}{2}$ $(k = \text{integer})$, then $2\lambda T \sim \chi^2_k$.

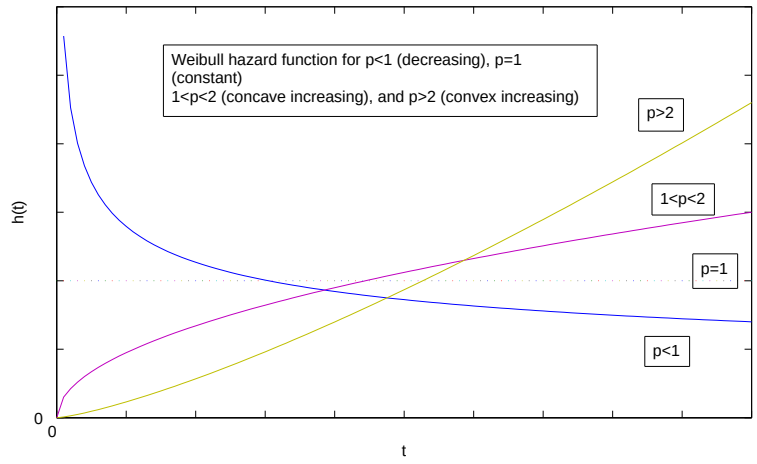
The following plot shows the shape of the Gamma hazard function for different values of the shape parameter α. The case α=1 corresponds to the exponential distribution (constant hazard function). When α is greater than 1, the hazard function is concave and increasing. When it is less than one, the hazard function is convex and decreasing.



Weibull Distribution:
The Weibull distribution can also be viewed as a generalization of the exponential distribution, and is denoted $W(\lambda, p)$. It is defined as follows:

$$
\begin{aligned}
F(t) &= 1 - e^{-(\lambda t)^p} \\
f(t) &= p\lambda^p t^{p-1} e^{-(\lambda t)^p} \\
h(t) &= p\lambda^p t^{p-1} \quad (\text{power of } t) \\
H(t) &= (\lambda t)^p.
\end{aligned}
\qquad
\begin{aligned}
t &> 0 \\
\lambda &> 0 \text{ (scale)} \\
p &> 0 \text{ (shape)}
\end{aligned}
$$

As shown in the following plot of its hazard function, the Weibull distribution reduces to the exponential distribution when the shape parameter p equals 1. When $p > 1$, the hazard function is increasing; when $p < 1$ it is decreasing.



The following properties of the Weibull distribution are easily verified.

- $T \sim W(\lambda, p),\ c > 0 \Longrightarrow cT \sim W(\frac{\lambda}{c}, p)$
- $T \sim W(\lambda, p) \Longrightarrow T^p \sim Exp(\lambda^p)$
- $W(\lambda, 1) \ = \ Exp(\lambda).$

Note: The Weibull distribution is sometimes parameterized as $H(t) = \lambda t^p$ instead of $H(t) = (\lambda t)^p$, in which case the expressions and properties above take on a somewhat different form.

1.3 Some Properties of Survival Time Random Variables

- $T_1,\ T_2, \ldots,\ T_n \quad i.i.d. \ \sim Exp(\lambda) \Longrightarrow T_1 + T_2 + \cdots + T_n \sim G(n, \lambda)$

  and $\ 2\lambda(T_1 + \cdots + T_n) \sim \chi^2_{2n}$

- Suppose $T_1, T_2, \cdots, T_n$ are i.i.d. $Exp(\lambda)$, and let $T_{(1)},\ T_{(2)}, \ldots,\ T_{(n)}$ denote the corresponding order statistics.

  For i=1,2,...,n, define

  $$Z_i = (n - i + 1)\left[T_{(i)} - T_{(i-1)}\right]$$

  where $T_{(0)} = 0$. That is,

  $$\begin{aligned} \text{i.e.,} \qquad Z_1 &= nT_{(1)} \\ Z_2 &= (n-1)[T_{(2)} - T_{(1)}] \\ &\vdots \\ Z_n &= T_{(n)} - T_{(n-1)}. \end{aligned}$$

  $Z_1, Z_2, \cdots, Z_n$ are sometimes called 'Normalized Spacings'. Imagine the 'window' in time extending from t=0 until t=$T_{(1)}$. The total amount of lifetime observed during this window is $nT_{(1)}$, since all n subjects are alive througout this time period. This is just Z_1. Next consider the 'window' extending from t=$T_{(1)}$ to $T_{(2)}$. The total observed time in this window is $(n-1)[T_{(2)} - T_{(1)}]$, since n-1 subjects survive through this window. This is just Z_2. Finally, the total observed time in the window extending from t=$T_{(n-1)}$ to $T_{(n)}$ is just $[T_{(n)} - T_{(n-1)}] = Z_n$, since only 1 subject passes through this window. The normalized spacings have an interpretation in terms of accumulated lifetime observed in specific cross-sectional views of the data. When the original T_i are i.i.d. exponential(λ) random variables, it can be shown that $Z_1, Z_2, \ldots, Z_n$ are also $i.i.d. \sim Exp(\lambda)$ (Exercise 4). That Z_1, which is n times the "gap" until the first failure, and Z_n, which is the gap between the next-to-last and last failure, have the same distribution speaks to the right tail of the exponential distribution.

- Poisson Process with parameter λ:

  $N(t)\ =\ \#$ events occuring in $(0, t)$ $\qquad\qquad \sim Pois(\lambda t)$.

Define $T_i =$ time between $(i-1)^{\text{st}}$ and i^{th} events, $i = 1, 2, \ldots$
(0^{th} event $=$ time 0).

Then $[N(t) = 0] \;=\; [T_1 > t] \Longrightarrow P(T_1 > t) = \bar{e}^{(\lambda t)} \frac{(\lambda t)^0}{0!} = \bar{e}^{\lambda t}$

i.e., $T_1 \sim Exp(\lambda)$.

It can also be shown that $T_1, T_2, \ldots$ are *i.i.d.* $Exp(\lambda)$.

- Suppose T is continuous and has cumulative hazard function $H(\cdot)$.

  Consider
  $$Y \stackrel{\text{def}}{=} H(T).$$
  Then
  $$Y \sim Exp(1) \qquad \text{(unit exponential!)}.$$
  See Exercise 5.

  Example 1: $T \sim W(\lambda, p) \Longrightarrow H(t) = (\lambda t)^p$.

  Thus, $\quad H(T) \equiv (\lambda T)^p \sim Exp(1)$.

  This result is analogous to what is sometimes called the "Probability Integral Transformation" for any continuous random variable. That is, if X has a continuous distribution and its c.d.f. is denoted $F(\cdot)$, then the random variable F(X) has the Uniform(0,1) distribution. As we will see, it is useful for generating realizations from specific survival distributions and for computing residuals from regression models for survival data.

1.4 Right Censored Observations

A common feature of survival (or time-to-event) data is the presence of right censored observations. We briefly review settings in which right-censored data can arise and introduce notation to distinguish the underlying T from what is actually observed.

Among the earliest forms of right censoring that were recognized and analyzed by statisticians arose in industrial life-testing settings, where the goal was to learn about the lifetime distribution of a manufactured item, and to find 'cost effective' ways of doing so. Two designs were commonly used, which we illustrate for the setting where the manufactured item is a light bulb and interest centers on the distribution function, say $F(\cdot)$, of the time until a bulb burns out:

Type I Censoring Suppose that we "plug in" n bulbs at time 0, and then observe them for c time units, noting the times until burn out for those that burn out by time c.

For the i^{th} bulb, let T_i = true lifetime.

Note that we observe T_i if and only if $T_i < c$;
otherwise, we know only that T_i exceeds c (right censored).

i.e., $T_1, \ldots, T_n$ *i.i.d.* $\sim F$,

but in general we observe only (U_i, δ_i) for $i = 1, 2, \ldots, n$

where $U_i = \min(T_i, c)$ = observed portion of T_i

and $\delta_i = 1(T_i \leq c)$ = censoring indicator.

e.g., Suppose $c = 10$:

Then if $T_i = 4$, we observe $(U_i, \delta_i) = (4, 1)$

But if $T_i = 17$, we would observe $(U_i, \delta_i) = (10, 0)$.

Thus, the 'data' from such a study would involve a (random) number, say r, of uncensored lifetimes, all of which are less than c, and $n - r$ censored observations at time c.

Type II Censoring Suppose instead that we "plug in" n bulbs, and then observe things until r (some pre-determined #) bulbs fail. Here we end up with r uncensored lifetimes and $n - r$ censored (at time c) lifetimes, but unlike Type I censoring, here r is a constant. Note that with this type of study, we are really observing the first r order statistics from a sample of size n.

i.e., $\quad T_1,\ T_2, \ldots,\ T_n\ \ i.i.d. \sim F$ (as with Type I) ,

and we observe

$$T_{(1,n)},\ T_{(2,n)}, \ldots,\ T_{(r,n)},$$

where $T_{(i,n)} = i^{th}$ smallest from among n times.

Note that we can also write the observation for bulb i in the form (U_i, δ_i). However, now U_i and δ_i depend on the values of the T_j for $j \neq i$. Thus, the n pairs $(U_1, \delta_1), (U_2, \delta_2), \cdots, (U_n, \delta_n)$ are dependent for Type II censoring, whereas they are independent for Type I censoring.

Random Censoring (commonly arising in biostatistics)
Define

$$c_1,\ c_2, \ldots,\ c_n \ \text{ constants}$$
$$T_1,\ T_2, \ldots,\ T_n \ \ i.i.d. \sim F.$$

Then suppose we observe $\quad (U_1, \delta_1), \ldots, (U_n, \delta_n),$

where

$$\begin{aligned} U_i &= \min(T_i, c_i) \\ \delta_i &= 1(T_i \leq c_i). \end{aligned}$$

For example, such censoring commonly arises in clinical trials where patients enroll sequentially at different points in calendar time, where 'survival' is measured from enrollment, and where the data are 'analyzed' at some later point, say $t*$, in calendar time. That is,

- T_i = time from entry into trial until relapse of disease for patient i, and
- c_i=time between enrollment of patient i and calendar time $t*$

As we will see later in the course, inferences usually proceed as if the c_i are known constants. However, for purposes of study design or when studying the asymptotic properties of estimators and tests, they are usually regarded as i.i.d. random variables (say, $C_1, C_2, \cdots, C_n$).

Key Assumption: Censoring is <u>noninformative</u>.

The large majority of statistical methods for failure time data assume that censoring acts 'noninformatively' of failure time. Loosely speaking, this means that being censored at time c tells us only that $T > c$. In terms of the potential censoring times C_i, noninformative censoring is achieved if C_i is independent of T_i, $i = 1, \cdots, n$; that is, if $C_i \perp T_i$. Mathematically weaker conditions have also been proposed and investigated (cf: Kalbfleisch & Prentice, p. 212-214).

Examples:

Censoring that is clearly informative: Suppose that T denotes time to death in a study and that subjects have an increased risk of 'dropping out' of the study (and thereby yielding a censored observation of survival time) if their disease status worsens. In the extreme case, subjects might drop out very soon prior to dying. In such cases, it is intuitively clear that a statistical method that assumes noninformative censoring might be severely biased and underestimate the true underlying hazard.

Censoring that is often noninformative: Suppose that survival time is censored because a subject in a study has not yet failed by the pre-scheduled date for the analysis of the study data. Here censoring is often noninformative.

Less clear situation: Suppose the survival times of subjects are censored if they move out of the study area and thereby can no longer be followed for survival. Can you envision specific circumstances when such censoring is and is not informative?

Exercises

Prove

1. $T \sim Exp(\lambda), c > 0 \Longrightarrow cT \sim Exp(\lambda/c)$

2. $T_1,\ T_2, \ldots,\ T_n\ \ i.i.d.\ Exp(\lambda) \Longrightarrow \sum_{i=1}^{n} T_i \sim G(n, \lambda)$

3. $T \sim W(\lambda, p) \Longrightarrow T^p \sim Exp(\lambda^p)$

4. $T_1, \ldots, T_n\ \ i.i.d.\ Exp(\lambda)$,
   $Z_i \stackrel{\text{def}}{=} (n - i + 1)[T_{(i)} - T_{(i-1)}] \Longrightarrow Z_i$ are $i.i.d.\ Exp(\lambda)$

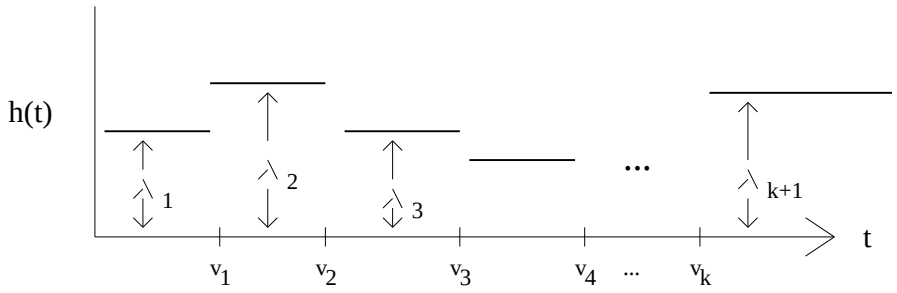
   HINT: Find distribution of $T_{(1)},\ T_{(2)}, \ldots,\ T_{(n)}$ and then consider Jacobian of transformation from these to $Z_1,\ Z_2, \ldots,\ Z_n$

5. Suppose that T is continuous distribution with cumulative hazard function $H(\cdot)$. Show that $H(T) \sim Exp(1)$.

6. A flexible parametric family is the **piecewise exponential**. For known constants $0 < v_1 < v_2 < \cdots < v_k$, the hazard function is given by

   $$h(t) = \lambda_j \quad \text{for } v_{j-1} \le t < v_j$$

   for $j = 1,\ 2, \ldots,\ k + 1$, where $v_o = 0$ and $v_{k+1} = \infty$.



Since any smooth function can be closely approximated by a step-function for sufficiently large k and choice of $v_1, \ldots, v_k$, it is clear that this family is quite flexible.

$\longrightarrow$ Find an expression for $f(t)$ and $S(t)$ for an arbitrary value of $t \, \varepsilon \, [v_{j-1}, v_j)$.

7. Suppose T has a discrete distribution, with p.m.f.

$$f(w_j) = P[T = w_j] = p_j \qquad j = 1, 2, \ldots,$$

where $0 \leq w_1 < w_2 < \cdots$ are known.

Then the hazard function can be defined as $h_j = P(T = w_j | T \geq w_j)$ for $j = 1, 2, \ldots$ Using this notation, show that the p.m.f. and survivor function for T can be expressed as

$$f(t) = \begin{cases} h_j \prod_{l=1}^{j-1}(1 - h_l) & \text{for } t = w_j \\ 0 & \text{otherwise} \end{cases}$$

and

$$S(t) = \prod_{\substack{j \text{ s.t.} \\ w_j \leq t}} (1 - h_j)$$

8. For a continuous survival time r.v. T, verify that

$$\begin{aligned} h(t) &= f(t)/S(t) \\ &= -\tfrac{d}{dt} ln[S(t)], \end{aligned}$$

that

$$S(t) = e^{-H(t)},$$

and that

$$f(t) = e^{-H(t)} \cdot h(t)$$

9. Suppose (T,C) denote a survival time and potential censoring time, respectively, and that Z is a covariate (such as a binary indicator of treatment group). Suppose that $T \perp C \mid Z$. Does it follow that $T \perp C$? Suppose on the other hand that $T \perp C$. Does it follow that $T \perp C \mid Z$? Prove the statements that are true and give specific examples if not.

10. Show that if k is an integer and $T \sim G(\frac{k}{2}, \lambda)$ then $2\lambda T \sim \chi^2_k$.

11. T time to death. Suppose survival times of subjects are censored if they move out of the study area and thereby can no longer be followed for survival. Give examples where one could envision that such censoring is and is not informative.

12. Show that $S_1(t) = [S_0(t)]^\theta$ is equivalent to $h_1(t)/h_0(t) = \theta$ (proportional hazards).

Additional Reading

For more motivating examples, see Cox & Oakes (1984) and Klein & Moeschberger (1997). If you are interested in learning more about early statistical methods for survival data that were motivated by industrial life-testing situations, see Lawless (2003). For a more technical discussion of informative censoring, see Lagakos (1979) and the text by Kalbfleisch & Prentice (2002). We will encounter the importance of noninformative censoring later in the course when we show that most of the key statistical methods for analyzing survival data can be expressed in terms of martingale processes.

References

Cox DR and Oakes D (1984). Analysis of Survival Data. Chapman and Hall, New York.

Kalbfleisch JD and Prentice RL (2002). The Statistical Analysis of Failure Time Data. Second Edition. Wiley, New York.

Klein and Moeschberger (1997). Survival Analysis: Techniques for Censored and Truncated Data. Springer-Verlag, New York.

Lagakos SW (1979). General right censoring and its impact on the analysis of survival data, Biometrics 35:139-156.

Lawless JF (2003). Statistical Models and Methods for Lifetime Data. Wiley, New York.