Chapter 1

Survival Analysis and Counting Processes

1.1 Introduction

1.1.1 Historical Remarks

The subject of survival analysis dates back to John Graunt [16] (1662) and Edmond Halley [17] (1693), who were the first to draw up life tables. Loosely, let $N_t$ denote the number of individuals of age $t$ and $n_t$ the number who survive to age $t$, but die before $t + 1$, where $t$ is an integer number of years. The probability that an individual survives to age $T$ may then be estimated by

$$P(\text{age} > T) = \prod_{t=0}^{T-1} \left(1 - \frac{n_t}{N_t}\right).$$

Life tables indicate the mortality rates for individuals of a certain age. They also indicate causes of death.

One of the first attempts at statistical inference was made by Daniel Bernoulli [8] (1760) where he estimated the increase in life expectancy if smallpox were eliminated. He based his calculations on the life tables of Edmond Halley. His calculations make a key assumption, the independence of competing risks. That is, he assumed that susceptibility to smallpox was independent of susceptibility to other possible causes of death. He estimated that life expectancy from birth was around 27 years and that eliminating smallpox would add 3 years to this figure.

The most important assumption for constructing a life table is that death rates do not change over time. If there is a trend, then the life table will be biased. From Bernoulli's day onwards, death rates have been decreasing in the Western world; this was the beginning of the demographic transition. Therefore, life tables based on this assumption have always understated life expectancy.

Survival analysis is a statistical methodology used in settings where one is interested in the occurrence of events, trying to understand the causes or establishing risk factors. A single event may be described by survival curves and hazard rates, its dependence on other factors (covariates) described by regression models. The connecting together of several events for an individual as they occur over time yields event histories.

The main difference between survival analysis and other branches of statistics is the form in which
the data comes. For many statistical situations, an experiment is carried out and the data collected and
analysed under the assumption that the observations are an observed random sample (i.e. observations
are realisations of independent identically distributed random variables). Survival analysis techniques
consider situations where such assumptions do not hold; the data comes as a mixture of complete and
incomplete observations. For example, consider a clinical study were 10 patients are observed over a
period to see if a certain event occurs (for example, relapse of some disease). The patients may enter
the study at different times. The event in question may occur while they are being observed, or they
may die of a different cause, or the study may be closed after a period of 12 years while some of the
patients are still alive and the event has not taken place.

1.1.2 Right Censoring, Left Truncation

Incomplete observations are termed censored survival times. When the data cut-off occurs on the right
hand side, the subject is lost to study from time a time \( \tau \) onwards but is known up to time \( \tau \), this is
known as right-censoring. A set of individuals for which the event of interest has not happened by time
\( t \) and who have not been censored by time \( t \) is known as the risk set at time \( t \). Censoring is broadly
characterised according to three types:

- **Type I censoring**: For type 1 censoring, individual \( i \) and the only piece of information recorded
  is whether \( \{ T_i > c_i \} \) (failure has not occurred by censoring time) or \( \{ T_i \leq c_i \} \) (failure occurs before,
  or on, censoring time).

- **Type II censoring**: The experiment is run until a fixed time \( t \). We observe \( r \) failures,
  \( T_{(1)}, \ldots, T_{(r)} \), the \( r \) failures that occurred up to \( t \); the only information for the remaining \( n - r \)
  individuals is that they survived for longer than \( t \).

- **Random censoring**: This is similar to Type I censoring, except that the \( c_i \) are observed values
  of random variables \( C_i \) which are independent of the survival times.

A concept related to right-censoring is left-truncation. This occurs if patients come under observa-
tion some time after an initiating event. For instance, in a study of myocardial infarction (the medical
term for 'heart attack'), only those who survive the initial phase and reach the hospital will be included
in the study. Time zero for an individual patient may be the time of infarction, and if the patient
reaches the hospital and is included in the study, this may be different for different patients. The term
delayed entry is also used.

1.1.3 Hazard Rates and Basic Survival Distributions

Let \( \tau \) be a positive random variable, the waiting time for failure. Suppose that \( \tau \) has a continuous
positive density \( f \). The distribution function is \( F(t) = \int_0^t f(s)ds \). The survival function is \( S(t) = 1 - F(t) = \mathbb{P}(T > t) \). The hazard rate is defined as:

\[
\alpha(t) = -\frac{d}{dt} \log S(t).
\] (1.1)
This may be written as:

\[ \alpha(t) = \frac{f(t)}{S(t)} = \lim_{h \to 0} \frac{1}{h} \mathbb{P}(t < T \leq t + h \mid T > t). \]

The intuition is that \( \alpha(t) \, dt \) represents the conditional probability of failing in the time interval \((t, t + dt]\) given survival until time \(t\). It follows from this that \( \int_0^\infty \alpha(t) \, dt = +\infty \) (left as an exercise). It also follows that if \( \tau_1, \ldots, \tau_n \) are \( n \) independent variables with hazard rates \( \alpha_1, \ldots, \alpha_n \) respectively, then \( \min_{i \in \{1, \ldots, n\}} \tau_i \) has hazard rate \( \sum_j \alpha_j \) (straightforward exercise).

Some of the basic survival distributions are now given:

1. **Exponential Distribution** The simplest continuous time survival distribution is the exponential distribution, \( T \sim \text{Exp}(\lambda) \). This satisfies the so-called *memoryless property*:

\[ \mathbb{P}(T > t + s \mid T > s) = \mathbb{P}(T > t) \]

and is the only continuous distribution which has this property. This models a situation where death occurs due to a single random, unexpected, serious disaster.

2. **Gamma Distribution** While the exponential distribution may model a situation where death occurs due to a single serious disaster, the Gamma distribution is the so-called *multiple hit model*. If \( T \sim \text{Gamma}(k, \lambda) \), then \( T \) is the sum of \( k \) independent \( \text{Exp}(\lambda) \) variables; the individual can sustain \( k - 1 \) hits and dies on hit number \( k \).

3. **Weibull Distribution** This is the distribution with survival function

\[ S(t) = \exp\{-\alpha t^b\} \quad t \geq 0 \]

for parameters \( a \) and \( b \). If \( U \sim \text{Unif}(0, 1) \), then it is straightforward to show that \( e^{-\log U}^d \) is Weibull with \( a = (1/c)^{1/d} \) and \( b = 1/d \). The Weibull introduces ageing; for \( b < 1 \), the hazard rate decreases with age, while for \( b > 1 \), the hazard rate increases with age.

**Independence of Competing Risks** Independence of competing risks means that the times until the various risks cause death are independent of each other. Suppose that there are \( n \) risks; death occurs as soon as any one of \( n \) possible events happens. Let \( T_j \) denote the time at which risk \( j \) causes death if the subject is not already dead, for \( j = 1, \ldots, n \). Then the time to death is \( T = \min(T_1, \ldots, T_n) \). The assumption of the independence of competing risks is that \( T_1, \ldots, T_n \) are independent random variables. Under this assumption, the hazards for each of the risks may be estimated by treating failure due to any other cause as censoring.

### 1.2 Survival Times and Counting Processes

Observational or experimental data gathered over time are often modelled by a stochastic process. Data counting the number of events of different types that occur over time are modelled with counting processes. The following discussion gives the basic probabilistic set up for the processes.
**Notation**  \( \Omega \) will be used to denote the space of outcomes of a random experiment, \( \omega \in \Omega \) a generic outcome, \( \mathcal{F} \) the event space and \( \mathbb{P} \) a probability measure over \( \mathcal{F} \).

Formally, \( \mathcal{F} \) is a \( \sigma \)-algebra; that is, it satisfies

1. \( \phi \in \mathcal{F} \) and \( \Omega \in \mathcal{F} \), where \( \phi \) is the empty set,
2. \( A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F} \),
3. If \( (A_j)_{j=1}^{\infty} \) is a collection of events in \( \mathcal{F} \) then \( \bigcup_{j=1}^{\infty} A_j \in \mathcal{F} \).

The probability \( \mathbb{P} \) gives a number between 0 and 1 for each \( A \in \mathcal{F} \) and satisfies the *Kolmogorov Axioms*

1. \( \mathbb{P}(A) \geq 0 \ \forall A \in \mathcal{F} \),
2. \( \mathbb{P}(\Omega) = 1 \),
3. If \( (A_j)_{j=1}^{\infty} \) are disjoint, then \( \mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j) \).

The formal definition of a random variable is as follows:

**Definition 1.1** (Random variable). A function \( Z : \Omega \to \mathbb{R} \) is a random variable or measurable relative to \( \mathcal{F} \) if

\[
\{ Z \leq x \} := \{ \omega : Z(\omega) \leq x \} \in \mathcal{F} \quad \forall x \in \mathbb{R}.
\]

That is, \( Z \) is a measurable mapping from \((\Omega, \mathcal{F}, \mathbb{P})\) to \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\).

**Definition 1.2** (Stochastic Process). A stochastic process is a family of random variables \( X = \{ X(t) : t \in \Gamma \} \) defined on a probability space \((\Omega, \mathcal{F}, \mathbb{P})\), where \( \Gamma \) is an indexing set.

The set \( \Gamma \) indexes time and is usually \( \mathbb{R}_+ \) or \( \mathbb{Z}_+ \); this course deals predominantly with continuous time; \( \mathbb{R}_+ \).

The event space \( \mathcal{F} \) is a collection of all the possible events (collections of outcomes) that are of interest to us. As time progresses, more information becomes available. Let \( \mathcal{F}_t \) denote the collection of events of interest, which can be determined up to time \( t \). For example, let \( T \) denote a survival time, then if we have been observing up to time \( t \), we know whether or not \( \{ T \leq s \} \) or \( \{ T > s \} \) for each \( s \in [0,t] \).

The \( \sigma \)-algebra of events known up to time \( t \) is the smallest \( \sigma \)-algebra containing all those events.

**Definition 1.3** (Filtration). 1. A family of sub-\( \sigma \)-algebras \( \{ \mathcal{F}_t : t \geq 0 \} \) of \( \mathcal{F} \) is increasing if

\[
s \leq t \Rightarrow \mathcal{F}_s \subseteq \mathcal{F}_t.
\]

An increasing sequence of sub-\( \sigma \)-algebras is called a filtration.

2. For a filtration \( \{ \mathcal{F}_t : t \geq 0 \} \), \( \mathcal{F}_{t+} = \cap_{s \geq 0} \mathcal{F}_{t+s} \) and \( \mathcal{F}_{t-} = \cup_{h \geq 0} \mathcal{F}_{t-h} \).

3. A filtration is right continuous if, for any \( t \), \( \mathcal{F}_{t+} = \mathcal{F}_t \).
4. A stochastic basis is a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) equipped with a right continuous filtration \(\{\mathcal{F}_t : t \geq 0\}\) and is denoted by \((\Omega, \mathcal{F}, (\mathcal{F}_t : t \geq 0), \mathbb{P})\).

5. A stochastic basis is complete if \(\mathcal{F}\) contains any subset of a \(\mathbb{P}\)-null set and for each \(t > 0\), \(\mathcal{F}_t\) contains all \(\mathbb{P}\)-null sets of \(\mathcal{F}\).

In future, we’ll assume that the event \(\sigma\)-algebras with which we are dealing are complete.

**Definition 1.4** (Adapted). A stochastic process \(X\) is adapted to a filtration \(\{\mathcal{F}_t : t \geq 0\}\) if for every \(t \geq 0\), \(X(t)\) is \(\mathcal{F}_t\) measurable.

**Heuristic** To say that \(X_t\) is \(\mathcal{F}_t\) measurable simply means that the information whether \(1_A(\omega) = 1\) or 0 for each \(A \in \mathcal{F}_t\) determines \(X(t, \omega)\). If \(X\) is adapted to \(\{\mathcal{F}_t : t \geq 0\}\), then the information whether \(1_A(\omega) = 1\) or 0 for each \(A \in \mathcal{F}_t\) determines the trajectory up to time \(t\). That is, it determines \(\{X(s, \omega) : 0 \leq s \leq t\}\).

The most natural filtrations are histories of stochastic processes. The natural filtration of a process \(X\) is the filtration such that \(\mathcal{F}_t\) represents the information generated by the process \(X\) on the interval \([0, t]\).

**Example 1.5.**

Let \(X_1, X_2, X_3, \ldots\) be independent identically distributed,

\[
\mathbb{P}(X_j = 1) = \mathbb{P}(X_j = -1) = \frac{1}{2}
\]

Let \(S_t = \sum_{j=1}^{t} X_j\) for \(t = 1, 2, 3\). This is a simple random walk. Let \(S\) denote the trajectory \((S_1, S_2, S_3)\).

Here, the trajectory \((S_1, S_2, S_3)\) determines \((X_1, X_2, X_3)\) and vice versa. Therefore, the probability space may be defined in terms of outcomes of \(X = (X_1, X_2, X_3)\). This is:

\[
\Omega = \{(1,1,1), (1,1,-1), (1,-1,1), (1,-1,-1), (-1,1,1), (-1,1,-1), (-1,-1,1), (-1,-1,-1)\}.
\]

For a discrete process, constructed from discrete and finite random variables, the representation of information through a \(\sigma\)-algebra may seem unnecessary. The filtration is as follows:

\[
\mathcal{F}_0 = \{\emptyset, \Omega\}
\]

\[
\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1.
\]
At time step 0, it is known that \(X \in \Omega\) and \(X \notin \emptyset\).

\[
\mathcal{F}_1 = \{\emptyset, \Omega, (1,1,1,1,1,-1,1,1,-1,-1), (1,1,1,1,-1,1,1,-1,-1), (1,1,1,1,1,1,-1,1,-1,-1), (1,1,1,1,1,1,1,1,-1,-1,-1)\}
\]

Here \(\mathcal{F}_1\) gives a complete listing of the sets \(A\) for which it is known at time 1 whether or not \(X \in A\). Similarly,
\[ F_2 = F_1 \cup \{(1,1,1), (1,1,-1)\}, \]
\[ \{(1,-1,1), (1,-1,-1)\}, \{(1,1,1), (-1,1,-1)\}, \{(1,-1,1), (-1,1,-1)\}\],
while \( F_3 = F \) is simply the set of all subsets of \( \Omega \).

A concept of importance is that of conditional expectation. Loosely, the conditional expectation \( E[X|F] \) where \( F \) is a \( \sigma \)-algebra of events is a random variable \( E[X|F](\omega) = E[X|B] \) where \( B \) is any set in \( F \) such that \( \omega \in B \). Formally, conditional expectation is defined as follows:

**Definition 1.6** (Conditional Expectation). Suppose \( Y \) is a random variable on a probability space \( (\Omega,F,\mathbb{P}) \) and that \( G \subseteq F \). Let \( X \) be a random variable satisfying:

1. \( X \) is \( G \)-measurable,
2. \( E[Y1_B] = E[X1_B] \) for all \( B \in G \).

Then \( X \) is the conditional expectation of \( Y \) given \( G \), written \( E[Y|G] \).

**Example 1.7** (Random Walk).

Let \( S \) denote the simple random walk described earlier with filtration \( (F_1,F_2,F_3) \) given earlier. Then \( E[S_3|F_2] = S_2 \). For example, consider \( \omega = (1,1,-1) \). Then the set in \( F_2 \) to be conditioned on is:
\[ A = \{(1,1,1), (1,1,-1)\} \]
\[ E[S_3|A] = E[S_3|X_1 = 1, X_2 = 1] \]
since \( A \) provides no restriction on \( X_3 \). Then
\[ E[S_3|X_1 = 1, X_2 = 1] = E[X_1 + X_2 + X_3|X_1 = 1, X_2 = 1] = 2 + E[X_3] = 2 = S_2. \]

The others are similar giving that in all cases \( E[S_3|F_2] = S_2 \).

**Definition 1.8** (Counting Process). A counting process is a stochastic process \( \{N_t : t \geq 0\} \) adapted to a filtration \( \{F_t : t \geq 0\} \) with \( N(0) = 0 \), where for each \( t < +\infty \), \( \mathbb{P}(N(t) < +\infty) = 1 \) with paths which are right continuous, piecewise constant, and only have jump discontinuities where the jumps are of size 1.

Right continuous, piecewise constant, jumps of size 1 simply means that for jump times \( 0 = t_0 < t_1 < t_2 < \ldots < t_n \), the sample path is:
\[ N(t) = j : t \in [t_j, t_{j+1}) \quad j = 1, \ldots, n-1 \]

For a counting process, with natural filtration \( \{F_t : t \geq 0\} \), the \( \sigma \)-algebra \( F_t \) is simply the information of how many jumps took place in the time interval \( [0,t] \) and the times of these jumps.
1.2. SURVIVAL TIMES AND COUNTING PROCESSES

**Definition 1.9** (Martingale). Let \( X = \{X(t) : t \geq 0\} \) be a right continuous stochastic process with left limits and \( \{\mathcal{F}_t : t \geq 0\} \) a filtration, defined on a common probability space. \( X \) is a martingale with respect to \( \{\mathcal{F}_t : t \geq 0\} \) if:

1. \( X \) is adapted to \( \{\mathcal{F}_t : t \geq 0\} \)
2. \( \mathbb{E} \left[ |X(t)| \right] < +\infty \ \forall t < +\infty \)
3. \( \mathbb{E} \left[ X(t + s) | \mathcal{F}_t \right] = X(t) \) \( \mathbb{P} \)-a.s. for all \( s \geq 0 \) and \( t \geq 0 \).

\( X \) is a submartingale if 3. is replaced by

\[
\mathbb{E} \left[ X(t + s) | \mathcal{F}_t \right] \geq X(t) \quad \mathbb{P} \text{-} a.s.
\]

It is a supermartingale if 3. is replaced by

\[
\mathbb{E} \left[ X(t + s) | \mathcal{F}_t \right] \leq X(t) \quad \mathbb{P} \text{-} a.s.
\]

**Censoring** A censoring time is a random time \( U \) after which an individual can no longer be monitored in a study. If \( T \leq U \), then the death is recorded; if \( T > U \), then the recorded information is that the individual survived beyond time \( U \).

A single censored observation provides the building block for the approach taken in this course to survival data. Let \( T \) be a survival time. Let \( U \) be a non-negative random time. Let \( X = \min(T, U) \) and \( \delta = 1_{\{T\leq U\}} \). This is the indicator function for an uncensored observation of \( T \). Let \( \{N(t) : t \geq 0\} \) be the counting process defined by:

\[
N(t) = 1_{\{X\leq \delta, \delta > 1\}} = \delta 1_{\{T\leq U\}}.
\]

\( N(t) = 1 \) if the event has happened by time \( t \) and it happened before the censoring time. \( N(t) = 0 \) either if the event did not happen by time \( t \), or the observation was censored before the event happened.

**Theorem 1.10.** Let \( T \) be an absolutely continuous failure time random variable and \( U \) a censoring time variable with an arbitrary distribution. Set \( X = \min(T, U) \) and \( \delta = 1_{\{T\leq U\}} \). Set

\[
\alpha^\#(t) := \begin{cases} 
\frac{-\mathbb{E}[\mathbb{P}(T > s, U \geq b)|X = z]}{\mathbb{P}(X > t)} & \mathbb{P}(X > t) > 0 \\
\mathbb{P}(X > t) = 0.
\end{cases}
\]

\[
N(t) = 1_{\{X\leq \delta, \delta > 1\}}, \quad N^U(t) = 1_{\{X\leq \delta, \delta = 0\}}, \quad \mathcal{F}_t = \sigma\{N(u), N^U(u) : 0 \leq u \leq t\}.
\]

Let \( M \) be defined by

\[
M(t) = N(t) - \int_0^t 1_{\{X>u\}} \beta(u) du
\]

where \( \beta \) is a deterministic function. Then \( M \) is an \( (\mathcal{F}_t)_{t \geq 0} \) martingale if and only if \( \beta(u) = \alpha^\#(u) \) for almost all \( u \).
Proof  Firstly, when \( \mathcal{F}_t \) is known, \( M \) is also known, so that \( M \) is adapted to \( \mathcal{F}_t \). Secondly, since \( N \) is a counting variable, taking values 0 or 1, it follows directly that \( \mathbb{E} |M(t)| < +\infty \) for all \( t < +\infty \). Thirdly,

\[
\mathbb{E} \left[ M(t + s) \mid \mathcal{F}_t \right] = \mathbb{E} \left[ N(t + s) - \int_0^{t+s} 1_{\{X > u\}} \beta(u) du \mid \mathcal{F}_t \right] \\
= M(t) + \mathbb{E} \left[ N(t + s) - N(t) \mid \mathcal{F}_t \right] - \mathbb{E} \left[ \int_t^{t+s} 1_{\{X > u\}} \beta(u) du \mid \mathcal{F}_t \right].
\]

Now

\[
\mathbb{E} \left[ N(t + s) - N(t) \mid \mathcal{F}_t \right] = \mathbb{E} \left[ 1_{\{t < X \leq t+s, \delta = 1\}} | N(u), N^U(u), 0 \leq u \leq t \right].
\]

If either \( N \) or \( N^U \) has a jump before time \( t \), then \( 1_{\{t < X \leq t+s, \delta = 1\}} = 0 \), so \( \mathbb{E} \left[ N(t + s) - N(t) \mid \{ X \leq t \} \right] = 0 \). For the set

\[
\{ X > t \} = \{ N(u) = N^U(u) = 0 \quad \forall u \in [0, t] \},
\]

set \( k = \mathbb{E} \left[ N(t + s) - N(t) \mid \{ X > t \} \right] \). Then

\[
\mathbb{E} \left[ N(t + s) - N(t) \mid \{ X > t \} \right] = \frac{\mathbb{P}(N(t + s) - N(t) = 1 \mid \{ X > t \})}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(t < X \leq t+s, \delta = 1)}{\mathbb{P}(X > t)}.
\]

Furthermore,

\[
\mathbb{E} \left[ \int_t^{t+s} 1_{\{X > u\}} \beta(u) du \mid \{ X > t \} \right] = \frac{1}{\mathbb{P}(X > t)} \mathbb{E} \left[ \int_t^{t+s} 1_{\{X > u\}} \beta(u) du 1_{\{X > t\}} \right] = \frac{1}{\mathbb{P}(X > t)} \mathbb{E} \left[ \int_t^{t+s} 1_{\{X > u\}} \beta(u) du \right] = \frac{1}{\mathbb{P}(X > t)} \int_t^{t+s} \mathbb{P}(X > u) \beta(u) du.
\]

It follows that \( M \) is a martingale if and only if for \( \mathbb{P}(X > t) \),

\[
\mathbb{P}(t < X \leq t+s, \delta = 1) = \int_t^{t+s} \mathbb{P}(X > u) \beta(u) du.
\]

But

\[
\mathbb{P}(t < X \leq t+s, \delta = 1) = \mathbb{P}(t < T \leq t+s, T \leq U) = \int_t^{t+s} \left( -\frac{\partial}{\partial v} \mathbb{P}(T \geq v, U \geq u) \bigg|_{v=x} \right) du = \int_t^{t+s} \mathbb{P}(X > u) \alpha \beta(u) du
\]

and the result follows. \( \square \)
1.2. SURVIVAL TIMES AND COUNTING PROCESSES

This clearly holds when \( T \perp U \). It may also hold in other situations, but it clearly does not hold in general. Where Bernoulli was studying the effects of smallpox, the censoring time \( U \) was death due to smallpox, \( T \) the time of death due to other causes. The assumption of the analysis was that \( T \perp U \), which did not take into account the possibility that a person susceptible to smallpox may be more susceptible to other fatal illnesses (i.e. the event that \( U \leq t \) denoting death by smallpox before time \( t \) might suggest a greater risk from other causes).

In problems of estimation, it is easier to work with the cumulative hazard rate. Let

\[
A(t) = \int_0^t \alpha(s)1_{\{X\geq s\}}ds, \quad A^\#(t) = \int_0^t \alpha^\#(s)1_{\{X\geq s\}}ds.
\]

**Example 1.11 (\( T \perp U \)).**

Suppose that \( (T, U) \) have a joint distribution described by:

\[\mathbb{P}(T > t, U > s) = \begin{cases} \exp \{-\lambda t - \mu s - \theta ts\} & t \geq 0, s \geq 0 \\ 0 & \text{other.} \end{cases}\]

Then

\[\mathbb{P}(T > t) = \exp \{-\lambda t\} \quad t \geq 0\]

so that the net hazard rate is: \( \alpha(t) = \lambda \). The crude hazard rate is:

\[\alpha^\#(t) = -\frac{\partial}{\partial u}\mathbb{P}(T \geq u, U \geq t)\big|_{u=t} = \lambda + \theta t.\]

\[\square\]

In clinical trials where censoring is caused by the choice of a data analysis time, dependence between \( T \) and \( U \) is unlikely. In an observational study, where censoring is caused by incomplete follow-up which occurs for reasons that are associated with the failure, then \( T \perp U \) and \( \alpha^\#(t) \) may differ from \( \alpha(t) \).