# Essentials of Survival Time Analysis

Thomas Hillen

## 1 Essentials of Survival-Time Analysis

Mathematical models in the form of ordinary differential equations (ODEs) often use simple rate transition terms such as $\dot{x} = -\mu x$. In this context people often refer to an exponential process, and it is sometimes not clear what they mean. We like to use this chapter to understand where such terms in differential equations come from and what is really behind the assumption of an *exponential distribution*. This will also clarify the relationship between rates and probabilities and highlight a common misunderstanding of this relation. Furthermore, the general framework developed here allows us to make a connection to delay differential equations[1].

### *1.1 Basic notations*

We are interested in individuals that can change their state. For example, susceptible individuals can get infected, prey individuals can be hunted, juvenile individuals can mature, mature individuals can reproduce, etc. We are interested in the expected time that an individual stays in a given state (i.e. time to get infected, time to get eaten, time to mature, etc.).

- Let $a$ be the time that an individual spends in a given state. The time $a$ is called the *soujourn time*, the *interevent time*, the *survival time*, or the ... ?

Name of First Author
Name, Address of Institute, e-mail: name@email.address

Name of Second Author
Name, Address of Institute e-mail: name@email.address

[1] This section is based on chapter 12 and 13 of H. Thieme, Mathematics in Population Biology, Princeton University Press [**?**]

- Let $F(a)$ denote the probability that an individual has not left the state before or at time $a$. Often $F(a)$ is simply called the survival probability, where here survival must be understood as survival in a given state until the individuum moves to the next state. We call $F$ the *soujourn function*, or *survival function* and we assume that $F(a)$ is non-increasing and $F(0) = 1$. If $\lim_{a\to\infty} F(a) = 0$ then each individual has to leave the state eventually. If $F(a) = 0$ for all $a > c$, then there is a maximum state duration time $c$ and all individuals will have left before time $c$.
- If $T$ denotes a random variable for the time to exit a given state, then

$$F(a) = P(T > a).$$

- The function $G(a) = 1 - F(a) = P(T \le a)$ denotes the probability to have left before time $a$.

## *1.2 Conditional Probabilities and Exit Rates*

We are interested in the conditional probability to still remain in the state for $h$ time units longer, given that the individual stayed already up to time $a$. This conditional probability is given by

$$F(h|a) = \frac{F(a+h)}{F(a)}.$$

The conditional probability to exit exactly between time $a$ and $a+h$, given that the individual was in the state at time $a$ is then

$$\frac{F(a) - F(a+h)}{F(a)} = 1 - F(h|a).$$

If $F$ is differentiable, then we define the *exit rate* as

$$\mu(a) = \lim_{h\to 0} \frac{F(a) - F(a+h)}{hF(a)} = -\frac{F'(a)}{F(a)}. \tag{1}$$

Note that since $F$ is non-increasing the rate $\mu(a)$ is non-negative. If $F$ is not differentiable, then we still use (1) with the distributional derivative of $F$.

**Example 1: exponential distribution** The first and most important example is the exponential distribution. I.e. we assume that the exit time is exponentially distributed and the soujourn function is given by

$$F(a) = e^{-\gamma a}.$$

In this case we can easily compute the rate (1) as

$$\mu(a) = \gamma,$$

and the conditional probability

$$F(h|a) = e^{-\gamma h} = F(h).$$

Hence the conditional probability of surviving $h$ time units longer is independent of the time spend in the state. In fact, the exponential distribution is the only distribution with this property:

**Theorem 1.** *(From Propostion 12.8 in [**?**]) $F(h|a)$ is independent of $a$ if and only if $F(a) = e^{-\gamma a}$, for some constant $\gamma \geq 0$.*

In case that the time increment $h = \Delta t$ is small, we can expand the exponential and find the probability of leaving the state in the interval $[t, t+\Delta t]$ as

$$G(\Delta t) = 1 - F(\Delta t|t) = 1 - e^{-\gamma \Delta t} \approx \gamma \Delta t + o(\Delta t).$$

In fact, the equation

$$G(\Delta t) = \gamma \Delta t \tag{2}$$

is often used to explain the relationship of a rate to a probability. We see here that this relationship is an approximation for small time increments $\Delta t$.

We can consider a similar expansion for general (differentiable) survival probabilities $F(a)$ as

$$G(\Delta t) = 1 - F(\Delta t|t) \approx F'(0|t)\Delta t + F''(0|t)\frac{(\Delta t)^2}{2} + h.o.t.$$

Let us take a brief look at the common use of the relation (2) as it is often found in the literature. Consider recovery from a disease and let us assume that 2 individuals recover per day. Then the probability to recover in one day is $G(1) = 2/20 = 1/10$. The corresponding rate, according to (2) is $\gamma = G(1)/1 = 1/10$. The rate here has units day$^{-1}$. The probability to recover in 1/2 a day equals $G(0.5) = 1/20$ and the corresponding rate is $\mu = (1/20)/(1/2) = 1/10$. Similarly, the probability to recover in 2 days is $G(2) = 4/20$ and the rate is $\mu = (4/20)/2 = 1/10$. We see that the rate remains constant, but the probability of change depends on the time interval chosen. It should be noted, though, that the rate has units day$^{-1}$ and if these units are changed, to weeks$^{-1}$ for example, then the rate changes as well.

Let us consider a simple probabilistic model for the recovery process. If $I(t)$ denotes a random variable for the number of infected individuals at time $t$, then

$$I(t+\Delta t) = I(t) - G(\Delta t)I(t).$$

We subract $I(t)$ and divide by $\Delta t$ to obtain

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

Passing to the limit $\Delta t \to 0$ we arrive at an ODE

$$\dot{I}(t) = -\mu I(t).$$

## 1.3 Age-Structured Models

In the general case we found the rate

$$\mu(a) = -\frac{F'(a)}{F(a)}.$$

In this case it can be shown that the population density satisfies an age structured model (see [**?**] for details). The *McKendrick* model describes the population density $u(t,a)$ of the number of individuals with state-age $a$ at time $t$:

$$\begin{aligned} u_t + u_a &= -\mu(a)u \\ u(t,0) &= B(t) \\ u(0,a) &= u_0(t), \end{aligned} \tag{3}$$

where $B(t)$ describes the individuals that enter the state with state-age 0. In addition to (3) we also assume that no individual stays forever, i.e. $u(t,\infty) = 0$.

The total state contents is then

$$N(t) := \int_0^\infty u(t,a)da.$$

**Example: Exponential exit times:** In the case of exponential exit times $F(a) = e^{-\gamma a}$ we find $\mu(a) = \gamma$ and we can integrate (3) with respect to $a$:

$$\int_0^\infty u_t da + \int_0^\infty u_a da = -\mu \int_0^\infty u da$$

which gives a linear birth-death ODE for $N$:

$$\dot{N} = B(t) - \mu N(t) \tag{4}$$

For the general case of $F(a)$ it was shown in Thieme [**?**], that we can derive also an equations for $N(t)$.

**Theorem 2.** *(Thieme [**?**]). Assume $B(t)$ is continuous and $F(a)$ is continuously differentiable with $F'(a) \le KF(a)$, then*

$$\dot{N}(t) = B(t) - C(t) \tag{5}$$

*with*

$$C(t) = \int_0^t \mu(a)B(t-a)F(a)da + \int_t^\infty \mu(a)F(a)\frac{u_0(a-t)}{F(a-t)}da.$$

*If $F$ is not differentiable, we can still write $C(t)$ using the integral over the measure $dF(a)$ as:*

$$C(t) = -\int_0^t B(t-a)dF(a) - \int_t^\infty \frac{u_0(a-t)}{F(a-t)}dF(a). \tag{6}$$

**Example 2: fixed stage duration**. Another interesting example is the case where individuals stay in the state for exactly $\tau$ time units and then they leave immediately. In that case

$$F(a) = \begin{cases} 1 & a \le \tau \\ 0 & a > \tau \end{cases}. \tag{7}$$

The soujourn time $F$ is not differentiable, but we can take the distributional derivative as

$$F'(a) = -\delta_\tau(a),$$

which means the measure in (6) is

$$dF(a) = -\delta_\tau(a)da$$

Then (6) becomes

$$\begin{aligned} C(t) &= \int_0^t B(t-a)\delta_\tau(a)da + \int_t^\infty \frac{u_0(a-t)}{F(a-t)}\delta_\tau(a)da \\ &= \begin{cases} B(t-\tau) & t > \tau \\ u_0(\tau - t) & t < \tau \end{cases}. \end{aligned}$$

This leads, for $t > \tau$ to a delay differential equation for $N$:

$$\dot{N}(t) = B(t) - B(t-\tau).$$

## 1.4 Summary of survival time analysis

- The time that individuals spend in a given state can have a general probability distribution $F(a)$.
- The most important case is the exponential distribution $F(a) = e^{-\gamma a}$. In this case the rate $\mu(a) = \gamma$ is constant and the conditional probabilities to live $h$ time units longer do no depend on the actual survival time $a$. Typical transition terms in differential equation models are based on the exponential distribution. The understanding that the transition probability in a small time interval $[t, t+\Delta t]$ is given by $P_{\Delta t} = \gamma\Delta t$ is in fact an approximation to the true value of

$$P_{\Delta t} = 1 - e^{-\gamma\Delta t}.$$

- A fixed stage duration (7) leads to delay differential equations.