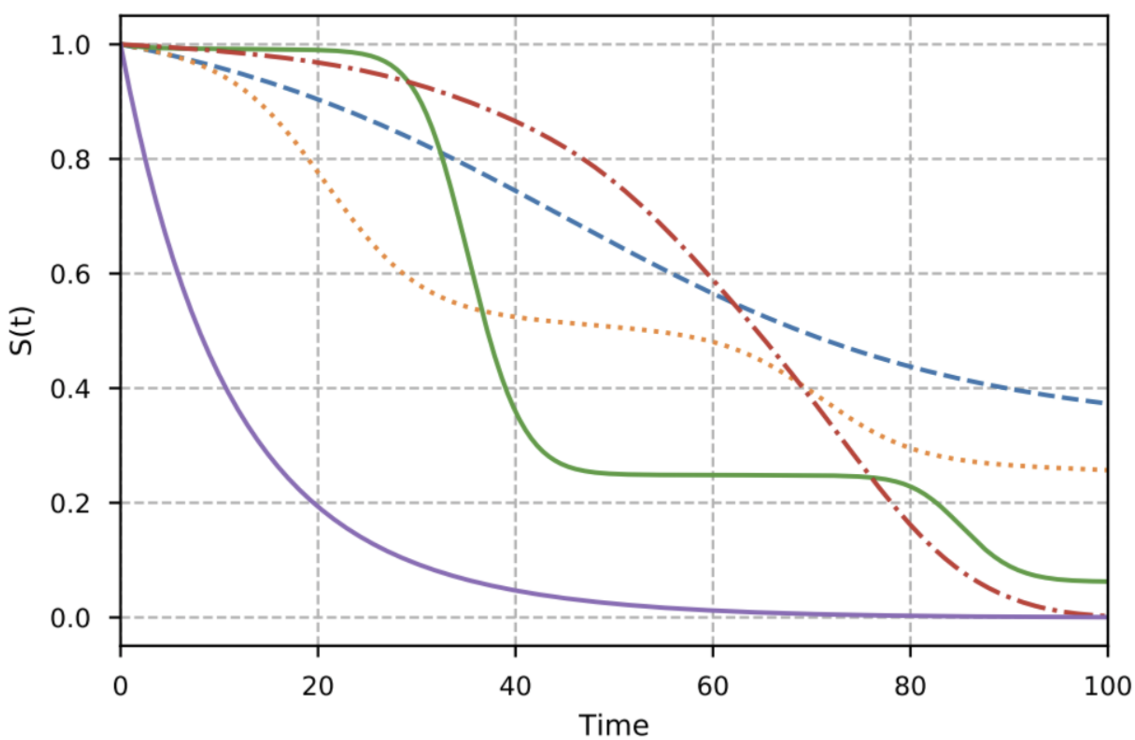
Alberto Andrés Valdés González.
Degree: Mathematical Engineer.
Work position: Data Scientist.
Mail: anvaldes@uc.cl/alberto.valdes.gonzalez.96@gmail.com
Location: Santiago, Chile.

Survival Analysis

Survival analysis is a branch of statistics for analyzing the expected duration of time until one event occurs, such as death in biological organisms and failure in mechanical systems. This topic is called reliability theory or reliability analysis in engineering, duration analysis or duration modelling in economics, and event history analysis in sociology.

Survival analysis attempts to answer certain questions, such as what is the **proportion of a population which will survive past a certain time?** Of those that survive, at what rate will they die or fail? Can multiple causes of death or failure be taken into account? How do particular circumstances or characteristics increase or decrease the probability of survival?



Math Review

We are working with a random variable T which represents the time of an events occurs. Of course $T > 0$.

Surival Function:

$$\boxed{S(t) = \mathbb{P}(T > t) = 1 - F(t)}$$

Non parametric maximum likelihood estimator (NPMLE) for survival function:

$$\boxed{\hat{S}(t) = \frac{1}{n} \cdot \sum_{i=1}^{n} I(T_i > t)}$$

Hazard Function:

$$\boxed{h(t) = \lim_{\epsilon \to 0^+} \frac{\mathbb{P}(t < T \leq t + \epsilon | T \geq t)}{\epsilon}}$$

Cumulative Hazard Function:

$$\boxed{H(t) = \int_0^t h(u) du}$$

Math Relations

$$\boxed{h(t) = \frac{f(t)}{S(t)}} \quad (1)$$

$$\boxed{H(t) = -ln(S(t))} \Rightarrow \boxed{S(t) = e^{-H(t)}} \quad (2)$$

$$\boxed{h(t) = f(t) \cdot e^{H(t)}} \quad (3)$$

Proof:

The (3) equation we can get it replacing (2) on (1). Because of that we only going to demostrate the (1) and (2) equation.

<u>(1):</u>

$$h(t) = \overset{\text{by definition}}{=} \lim_{\epsilon \to 0^+} \frac{\mathbb{P}(t < T \leq t + \epsilon | T \geq t)}{\epsilon} \overset{\text{by definition of conditional probability}}{=}$$

$$\lim_{\epsilon \to 0^+} \frac{\mathbb{P}(t < T \leq t + \epsilon \wedge T \geq t)}{\mathbb{P}(T \geq t)} \cdot \frac{1}{\epsilon}$$

Now we have that:

$$\{t < T \leq t + \epsilon \wedge T \geq t\} = \{t \leq T \leq t + \epsilon\}$$

$\Rightarrow$

$$h(t) = \lim_{\epsilon \to 0^+} \frac{\mathbb{P}(t < T \leq t + \epsilon \wedge T \geq t)}{\mathbb{P}(T \geq t)} \cdot \frac{1}{\epsilon} \overset{\text{by previous equation}}{=}$$

$$\lim_{\epsilon \to 0^+} \frac{\mathbb{P}(t \leq T \leq t + \epsilon)}{\mathbb{P}(T \geq t)} \cdot \frac{1}{\epsilon} \overset{\text{by definition of } S(t)}{=} \lim_{\epsilon \to 0^+} \frac{\mathbb{P}(t \leq T \leq t + \epsilon)}{S(t)} \cdot \frac{1}{\epsilon}$$

$$= \frac{1}{S(t)} \cdot \lim_{\epsilon \to 0^+} \frac{\mathbb{P}(t \leq T \leq t + \epsilon)}{\epsilon} = \frac{1}{S(t)} \cdot \lim_{\epsilon \to 0^+} \frac{F(t+\epsilon) - F(t)}{\epsilon} \overset{\text{by definition of derivative}}{=}$$

$$\frac{1}{S(t)} \cdot F'(t) = \frac{f(t)}{S(t)}$$

$\Rightarrow$

$$\boxed{h(t) = \frac{f(t)}{S(t)}}$$

<u>(2):</u>

$$H(t) = \int_0^t h(u) du \overset{\text{by (1)}}{=} \int_0^t \frac{f(u)}{S(u)} du \overset{\text{by (1)}}{=}$$

Create the variable $z = S(u) = 1 - F(u) \;\Rightarrow\; dz = -f(u) du$

$\Rightarrow$

$$H(t) = \int_0^t \frac{f(u)}{S(u)} du \overset{\text{by substitution}}{=} \int_{S(0)=1}^{S(t)} \frac{-\,dz}{z} = -ln(|S(t)|) + ln(|1|) = -ln(|S(t)|) \overset{S(t)>0}{=} -ln(S(t))$$

$\Rightarrow$

$$\boxed{H(t) = -ln(S(t))}$$

Exponential Distribution

Consider $T \sim Exp(\lambda)$.

As we know $f(t) = \lambda \cdot e^{-\lambda \cdot t}$, $F(t) = 1 - e^{-\lambda \cdot t}$. Now we have to compute $S(t)$, $h(t)$ and $H(t)$.

First we going to compute $S(t)$:

$S(t) = 1 - F(t) = 1 - [1 - e^{-\lambda \cdot t}] = e^{-\lambda \cdot t}$

$\Rightarrow$

$$\boxed{S(t) = e^{-\lambda \cdot t}}$$

Second we going to compute $h(t)$:

$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda \cdot e^{-\lambda \cdot t}}{e^{-\lambda \cdot t}} = \lambda$

$\Rightarrow$

$$\boxed{h(t) = \lambda}$$

Third we going to compute $H(t)$:

$H(t) = -ln(S(t)) = -ln(e^{-\lambda \cdot t}) = -(-\lambda \cdot t) = \lambda \cdot t$

$\Rightarrow$

$$\boxed{H(t) = \lambda \cdot t}$$
