Hazard rate, cumulative hazard and survival function in the absolutely continuous case

Let T be a lifetime with an absolutely continuous distribution. Then

$$
\begin{aligned}
\alpha(t) &= \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} \\
A(t) &= \int_0^t \alpha(s) ds \\
S(t) &= e^{-A(t)}
\end{aligned}
$$

Hazard rate and cumulative hazard in the general case

A general definition of hazard and cumulative hazard were given in Slides 8, starting from a general $S(t) = P(T > t)$ only assuming

- $S(t)$ is continuous from the right
- $S(t)$ is non-increasing
- $0 < S(0) \leq 1$

Then it was shown that

$$dA(t) = P(t \leq T < t + dt | T \geq t) = -\frac{dS(t)}{S(t-)}$$

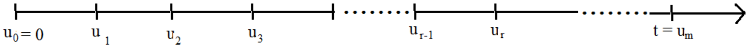
leading to the general expression

$$A(t) = -\int_0^t \frac{dS(u)}{S(u-)}$$

As a check, for the absolutely continuous case we would then have

$$A(t) = -\int_0^t \frac{dS(u)}{S(u-)} = \int_0^t \frac{-S'(u)du}{S(u)} = \int_0^t \frac{f(u)du}{S(u)} = \int_0^t \alpha(u)du$$

The discrete hazard and cumulative hazard from scratch



Let $u_0 = 0, u_1, u_2, \ldots$ be the values of T.

Reconsider $dA(t) = P(t \leq T \leq t + dt | T \geq t)$. Looking at the above axis it is clear that this is 0 unless t is one of the u_j.

For $t = u_j$, we get on the other hand,

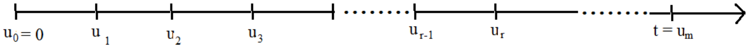
$$dA(t) = P(t \leq T < t + dt | T \geq t) = P(T = u_j | T \geq u_j) \equiv \alpha_{u_j}$$

which we called the *discrete hazard rate*.

Thus $A(t) = \int_0^t dA(u)$ is a function which jumps by α_{u_j} when $t = u_j$, so we get the natural result that

$$A(t) = \sum_{u_j \leq t} \alpha_{u_j}$$

The discrete survival function and the product-integral



For $u_m \leq t < u_{m+1}$ we now have

$$
\begin{aligned}
S(t) &= P(T > t) = P(T > u_m) \\
&= \prod_{j=1}^{m} P(T > u_j | T > u_{j-1}) = \prod_{j=1}^{m} (1 - P(T = u_j | T > u_{j-1})) \\
&= \prod_{j=1}^{m} (1 - \alpha_{u_j}) = \prod_{j=1}^{m} (1 - dA(u_j)) = \prod_{u_j \leq t} (1 - dA(u_j)) \quad (^*)
\end{aligned}
$$

This is a special case of the so called **product-integral**

$$S(t) = \prod\nolimits_{0 \leq u \leq t} (1 - dA(u))$$

When $A(t)$ makes jumps at discrete time points u_j, this is exactly given as in $(^*)$.

The general product-integral

Recall the general expressions

$$dA(u) = -\frac{dS(u)}{S(u-)}, \quad A(t) = -\int_0^t \frac{dS(u)}{S(u-)}$$

Solving $S(t)$ from these equations in the general case, leads to the so-called **product-integral** ,

$$S(t) = \prod\nolimits_{0 \le u \le t} (1 - dA(u))$$

which can be defined as a limit as follows:

Let $t > 0$ be fixed, and let $0 = u_0 < u_1 < u_2 < \ldots < u_m \equiv t$ define a partition of $[0, t]$. Let $m \to \infty$ in a way such that the spacings $u_j - u_{j-1}$ tend to 0. Then

$$S(t) = \prod\nolimits_{0 \le u \le t} (1 - dA(u)) =_{def} \lim_{m \to \infty} \prod_{j=1}^{m} [1 - (A(u_j) - A(u_{j-1}))]$$

A quick check for the continuous case

Suppose $A(t) = \int_0^t \alpha(u)du$ so that $A'(t) = \alpha(t)$.

Then for a given partition with large m and fixed $t = u_m$,

$$
\begin{aligned}
& \prod_{j=1}^{m}[1 - (A(u_j) - A(u_{j-1}))] \approx \prod_{j=1}^{m}(1 - \alpha(u_j)(u_j - u_{j-1})) \\
\approx \quad & \prod_{j=1}^{m} e^{-\alpha(u_j)(u_j - u_{j-1})} = e^{-\sum_{j=1}^{m} \alpha(u_j)(u_j - u_{j-1})} \\
\approx \quad & e^{-\int_0^t \alpha(u)du} \equiv S(t)
\end{aligned}
$$

Thus, the product-integral gives the correct result for continuous distribution, as we saw that it did for the discrete case.

The Kaplan-Meier estimator

Let's go back to the discrete survival function,

$$S(t) = \prod_{u_j \le t} (1 - \alpha_{u_j}) = \prod_{u_j \le t} (1 - dA(u_j))$$

An estimator of $S(t)$ can hence be obtained by putting an estimate for the function $A(t)$ in the above. This may be done by the *Nelson-Aalen estimator* and turns out to lead to the *Kaplan-Meier estimator*.

Recall the NA-estimator $\hat{A}(t) = \int_0^t \frac{dN(s)}{Y(s)}$. This is a discrete function with jumps $\frac{dN(s)}{Y(s)}$, i.e., it jumps at the failure times T_i. Thus we can write

$$\hat{S}(t) = \prod_{0 \le u \le t} (1 - d\hat{A}(u)) = \prod_{0 \le u \le t} \left(1 - \frac{dN(u)}{Y(u)}\right) = \prod_{T_i \le t} \left(1 - \frac{1}{Y(T_i)}\right)$$