Explaining $S(t) = \exp(-H(t))$

Notation

Notation: we use T to denote the random event time, $S(t) = P(T > t)$ to denote the survival function, and

$$h(t) = \lim_{\Delta t \to 0} P(t < T < t + \Delta t \,|\, T > t)/\Delta t, \qquad H(t) = \int_0^t h(s)ds$$

to denote the hazard and cumulative hazard function, respectively. The assumption is that the underlying event time distribution is continuous, so that we don't have to worry about differences between $<$ vs $\leq$ or $>$ vs $\geq$.

Explanation

The way to think about $P(T > t)$ (this will also be useful when we learn about the Kaplan-Meier estimate next week) is to divide the interval $(0, t)$ into many small intervals. Let's take them all of the same length, Δt, and let Δt get smaller and smaller. You remember from probability theory that $P(B) = P(B \,|\, A)P(A)$. We use that to say that

$$P(T > 2\Delta t) = P(T > 2\Delta t \,|\, T > \Delta t) \cdot P(T > \Delta t) \tag{1}$$

and, using (1)

$$\begin{aligned} P(T > 3\Delta t) &= P(T > 3\Delta t \,|\, T > 2\Delta t) \cdot P(T > 2\Delta t) = \\ &= P(T > 3\Delta t \,|\, T > 2\Delta t) \cdot P(T > 2\Delta t \,|\, T > \Delta t) \cdot P(T > \Delta t). \end{aligned}$$

The next step is that we write each of the conditional probabilities in the previous equation as one minus the probability of its complement, so

$$\begin{aligned} P(T > 3\Delta t) &= (1 - P(2\Delta t < T < 3\Delta t \,|\, T > 2\Delta t)) \cdot \\ &\quad \cdot (1 - P(\Delta t < T < 2\Delta t \,|\, T > \Delta t)) \\ &\quad \cdot (1 - P(T < \Delta t)). \end{aligned}$$

Now note that each of the conditional probabilities in the previous equation resembles the definition of the hazard, except for the Δt factor. Since the hazard is the limit for Δt going to 0, and in our equations Δt is small, we have

$$P(T > 3\Delta t) \approx (1 - h(2\Delta t)\Delta t)(1 - h(\Delta t)\Delta t)(1 - h(0)\Delta t).$$

For general t we have

$$P(T > t) \approx (1 - h(0)\Delta t)(1 - h(\Delta t)\Delta t) \cdots (1 - h(t)\Delta t) = \prod_s (1 - h(s)\Delta t),$$

where the product is over a grid of time points of distance Δt from 0 to t. When you take the limit of Δt going to 0, this product becomes a product integral, which is used in many mathematical texts about survival analysis. We are going to use the approximation coming from the first order Taylor expansion $e^{-x} \approx 1 - x$, for x small, the next (ignored) term in the Taylor expansion being $+x^2/2$. Replacing each of the $(1 - h(s)\Delta t)$ terms by $e^{-h(s)\Delta t}$ (previous Taylor expansion), we get

$$P(T > t) \approx \prod_s (1 - h(s)\Delta t) \approx \prod_s \exp(-h(s)\Delta t) = \exp(-\sum_s h(s)\Delta t),$$

where in the last equation we use that the exponent of a sum is the product of it exponents. Finally, we recognize that the sum of terms which get increasingly smaller (as Δt goes to 0) becomes an integral. Moreover, each of the approximations actually becomes equality when $\Delta t \to 0$. This, in the end, leads to

$$P(T > t) = \exp(-\int_0^t h(s)ds).$$