PAGE 2024 DeepPumas Workshop

Survival Models

Mohamed Tarek

Senior Product Engineer at PumasAI Inc.

Research Affiliate at University of Sydney Business School

Time to event models

Definitions

- Instantaneous hazard

$$\lambda(t) > 0$$

- Cumulative hazard

$$\Lambda(t) = \int_0^t \lambda(t')\, dt'$$

- Survival function: probability of survival up to time t

$$S(t) = \exp(-\Lambda(t))$$

Definitions

- Failure function: probability of death/failure before time t

$$F(t) = 1 - S(t)$$

- Probability density function of time of death t

$$f(t) = \frac{dF}{dt} = \lambda(t) \cdot \exp(-\Lambda(t))$$

- Expected time of death $E[t]$

$$E[t] = \int_0^\infty t \cdot f(t)\, dt = \int_0^\infty S(t)\, dt$$

Log likelihood

The log likelihood for censored survival data is given by the following 2 formulas:

- For censored subjects at time t (patient survived until time t)

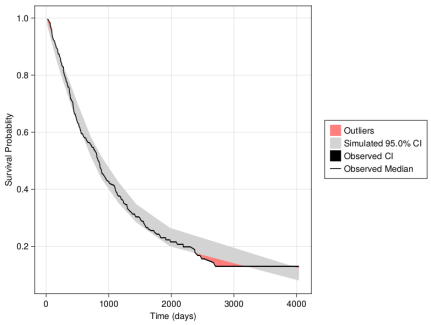
$$\log \text{likelihood} = \log S(t) = -\Lambda(t)$$

- For subjects dead at time t

$$\log \text{likelihood} = \log f(t) = \log \lambda(t) - \Lambda(t)$$

Time to event VPC procedure

Example



Time to event VPC procedure

1. Simulate a synthetic population a given number of samples (`samples`, default 499). For each subject:
    1. Evaluate the cumulative hazard function Λ at `nT` (default 10) time points between `minT` and `maxT`.
    2. Use a cubic spline to interpolate between the Λ values.
    3. Use inverse CDF transform sampling to sample the time of death from the cumulative hazard function.

Time to event VPC procedure

2. Stratify the observed and simulated populations by the stratification variable.
3. For each simulated population stratum:
   1. Estimate the Kaplan Meier (KM) curve. d_i is the number of deaths at t_i and n_i is the number of people at risk at time t_i.

$$\hat{S}(t) = \prod_{i:t_i<t} \left(1 - \frac{d_i}{n_i}\right)$$

   2. Combine all simulated populations' KM curves into one data frame.
   3. Do quantile regression with smoothing to get smooth curves for the quantiles at a number of nodes `nnodes` (default 11).

Time to event VPC procedure

4. For each observed population stratum, estimate the KM curve.
5. Plot the observed KM curve against the smoothed quantiles for each stratum.

Inverse CDF sampling

- If $R \sim \text{Uniform}(0, 1)$, then $-\log(1 - R) \sim \text{Exponential}(1)$.

$$\begin{aligned} F(t) &\leq R \\ 1 - S(t) &\leq R \\ \exp(-\Lambda(t)) &\geq 1 - R \\ \Lambda(t) &\leq -\log(1 - R) \end{aligned}$$

- The sample t is obtained using a root finding algorithm to find the root for $\Lambda(t) = -\log(1 - R)$.