Biostatistics (MATH11230)
Parametric estimators of the survival function

Vanda Inácio

University of Edinburgh

Semester 1, 2022/2023

Parametric estimation

General context

- The Kaplan–Meier and the Nelson–Aalen estimators of the survival function are nonparametric: besides independence between observations, we did not make any assumptions about the distribution of the survival times.
- In a parametric setup we instead choose a (parametric) distribution that we believe to appropriately describe the distribution of the survival times.
- As we have discussed last week, popular approaches are the Weibull, lognormal, and gamma distributions. But there are many more (e.g., Rayleigh, Pareto, etc).
- The exponential distribution, although analytically tractable, has the caveat of leading to a constant hazard function with respect to time and so it may be too simplistic in most situations.

Parametric estimation

General context

- Suppose we have a random sample of n individuals from a specific population whose true survival/event times are $T_1, \ldots, T_n$.
- However, due to right censoring (study period ends, dropouts/losses to follow up, competing risks, etc) we do not always have the opportunity of observing those survival times.
- Let us denote by C the censoring process and by $C_1, \ldots, C_n$ the (potential) censoring times for each individual.
- If a subject is not censored we have observed their event time (in this case, we may not observe the censoring time for this individual).
- On the other hand, if the subject is censored, we observe their censoring time (true event time is larger than censoring time).

Parametric estimation

Likelihood construction

- Let
$$Y_i = \min\{T_i, C_i\}, \qquad \text{and} \qquad \Delta_i = I(T_i \leq C_i).$$
- The observed data are then composed of couples (y_i, δ_i), for $i = 1, \ldots, n$.
- If y_i is not censored, observation i contributes with a factor to the likelihood equal to the density function of T evaluated at y_i, $f_T(y_i)$.
- If y_i instead represents a censored time, it is only known that t_i (true event time) exceeds y_i (observed event time) and the contribution to the likelihood is the probability that $T_i > c_i$ (equal to $\Pr(T_i > y_i)$) and this probability is $S_T(y_i)$, where S is the survival function of T.

Parametric estimation

Likelihood construction

- Assuming independent censoring and independent observations, we have that the likelihood is given by

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \left\{ f(y_i;\theta)^{\delta_i} S(y_i;\theta)^{1-\delta_i} \right\} \\
&= \prod_{i=1}^{n} \left\{ [h(y_i;\theta)S(y_i;\theta)]^{\delta_i} S(y_i;\theta)^{1-\delta_i} \right\} \\
&= \prod_{i=1}^{n} \left\{ h(y_i;\theta)^{\delta_i} S(y_i;\theta) \right\}.
\end{aligned}
$$

- Based on this likelihood, we can then determine the maximum likelihood estimate of θ (which may be a scalar or a vector) and having this we have all we need to estimate the survival, hazard, and cumulative hazard functions.

Parametric estimation

Example: exponential distribution

↪ Let us assume that the true survival times $T_1, \ldots, T_n$ follow an exponential distribution.

↪ We have that

$$f(t;\theta) = \theta e^{-\theta t}, \quad S(t;\theta) = e^{-\theta t}, \quad h(t;\theta) = \theta, \quad \theta > 0.$$

↪ According to the result in the previous slide we have that

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} h(y_i;\theta)^{\delta_i} S(y_i;\theta) \\ &= \prod_{i=1}^{n} \theta^{\delta_i} e^{-\theta y_i} \\ &= \theta^{\sum_{i=1}^{n} \delta_i} e^{-\theta \sum_{i=1}^{n} y_i} \end{aligned}$$

Parametric estimation

Example: exponential distribution

- ↪ The corresponding log likelihood is given by

$$\log L(\theta) = \log \theta \sum_{i=1}^{n} \delta_i - \theta \sum_{i=1}^{n} y_i,$$

and the derivative with respect to θ is

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log L(\theta) = \frac{1}{\theta} \sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n} y_i.$$

- ↪ Setting the derivative to zero and solving for θ:

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \log L(\theta) = 0 \Rightarrow \frac{1}{\theta} \sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n} y_i = 0$$

$$\Rightarrow \widehat{\theta} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} y_i}$$

Parametric estimation

Example: exponential distribution

- We take the second derivative to verify this is, in fact, the maximiser

$$\frac{d^2}{d\theta^2} \log L(\theta) = -\frac{1}{\theta^2} \sum_{i=1}^{n} \delta_i.$$

- Evaluating this second derivative at $\widehat{\theta}$ we conclude that it is negative and therefore $\widehat{\theta}$ is the maximum likelihood estimate.

- Note that in the case where all observations are uncensored, $\sum_{i=1}^{n} \delta_i = n$, and we recover the 'usual' maximum likelihood estimate, $\widehat{\theta} = n / \sum_{i=1}^{n} t_i$.

Parametric estimation

Example: Exponential distribution

- As an example let us consider the data we have used yesterday to manually compute the Kaplan–Meier estimate:

$$1 \quad 3 \quad 3 \quad 6^+ \quad 8^+ \quad 9 \quad 10^+$$

- We have four uncensored observations and so we know that $\sum_{i=1}^{7} \delta_i = 4$.

- We further have that $\sum_{i=1}^{7} y_i = 1 + 3 + 3 + 6 + 8 + 9 + 10 = 40$.

- Therefore, $\widehat{\theta} = 4/40 = 0.1$.

Parametric estimation

Example: Weibull distribution

- A more flexible alternative to the exponential distribution is the Weibull distribution.
- The defining feature of the Weibull distribution is that the corresponding hazard function is monotonic. It can be monotonically increasing, monotonically decreasing or constant with respect to time.
- It is fair to mention that the assumption of ever increasing or ever decreasing hazards may be unrealistic in certain situations as well.
- Nevertheless, the Weibull distribution is ubiquitous in survival analysis and we should consider it.

Parametric estimation

Example: Weibull distribution

↪ The probability density, survival and hazard functions of the Weibull distribution are given by

$$f(t;\lambda,\alpha)=\lambda\alpha t^{\alpha-1}e^{-\lambda t^{\alpha}},\quad S(t;\lambda,\alpha)=e^{-\lambda t^{\alpha}},\quad h(t;\lambda,\alpha)=\lambda\alpha t^{\alpha-1},\quad \lambda>0,\quad \alpha>0.$$

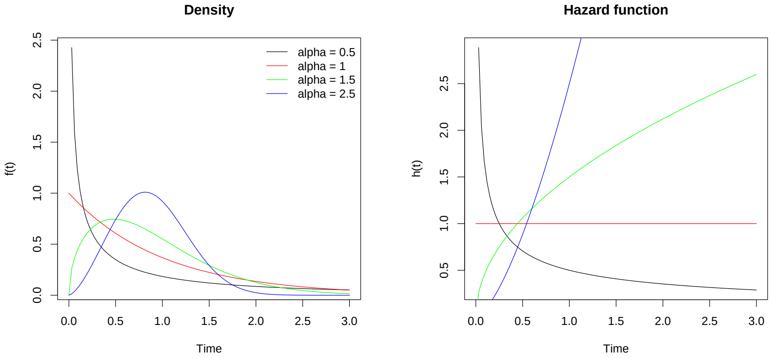
↪ Here λ is the scale parameter and α is the shape parameter. For $\alpha<1$, the hazard is decreasing, $\alpha=1$ the hazard is constant (see below), and $\alpha>1$ the hazard is increasing. The parameter scale λ stretches or squeezes the distribution.

↪ Note that the exponential distribution is a particular case of the Weibull distribution when the shape parameter α is equal to 1.

Parametric estimation

Example: Weibull distribution

$\hookrightarrow$ In the plots below $\lambda = 1$.



Parametric estimation

Example: Weibull distribution

- $\hookrightarrow$ The likelihood function of $\theta = (\lambda, \alpha)$ is given by

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} \left\{ h(y_i;\theta)^{\delta_i} S(y_i;\theta) \right\} \\
&= \prod_{i=1}^{n} \left\{ [\lambda\alpha y_i^{\alpha-1}]^{\delta_i} e^{-\lambda y_i^{\alpha}} \right\}.
\end{aligned}
$$

- $\hookrightarrow$ However, there is no closed form for the MLEs of $\theta = (\lambda, \alpha)$.

- $\hookrightarrow$ We can use, for instance, `optim` to maximise the above (log) likelihood function. We can also use, and it is indeed what we will do, the function `survreg` from the package `survival`.

- $\hookrightarrow$ Please see the Supplementary Materials file for more details about the fitting of the model.

Parametric estimation

- ↪ Before fitting a parametric model, we should check that it fits appropriately to the data at hand.
- ↪ For example, we could plot the Kaplan-Meier estimate of the survival function and the estimated survival model under an exponential or Weibull model.
- ↪ Personally, I prefer the Kaplan-Meier estimator which makes no assumptions regarding the shape of the distribution of the event times.
- ↪ In most medical applications, the default is to go with the Kaplan-Meier estimator.