SURVIVAL ANALYSIS

PART V, HY 2019

Additional remarks

- Poisson regression vs. exponential regression
- Late entry and left truncation
- Other patterns of incomplete survival data
- Estimation under censoring – a simple example on exponential regression

Poisson regression vs. exponential regression

- Survival data (t_i, d_i), $i = 1, \ldots, n$
- It was argued that the likelihood expressions for the constant hazard parameter under the exponential and Poisson models are equivalent
  - In the exponential regression model, the data include the times of failure or censoring $(\{t_i\})$ *and* the event indicators $(\{d_i\})$
  - In Poisson regression, the data include the event indicators $(\{d_i\})$, *given* the person-times $(\{t_i\})$
- If there were no covariates, the likelihood function of the hazard λ is

$$
\begin{aligned}
L(\lambda) &= \prod_{i=1}^{n} \mathrm{Exp}(t_i; \lambda) = \left(\prod_{i=1}^{n} \lambda^{d_i}\right) \times e^{-\lambda \sum_{i=1}^{n} t_i} = \lambda^{D} e^{-\lambda \sum_{i=1}^{n} t_i} \\
&\propto \frac{(\lambda \sum_{i=1}^{n} t_i)^{D} \exp(-\lambda \sum_{i=1}^{n} t_i)}{D!} = \mathrm{Poisson}(D; \lambda \textstyle\sum_{i=1}^{n} t_i)
\end{aligned}
$$

Poisson regression vs. exponential regression (2)

- Then consider adding covariates to the model, i.e. setting up an actual regression model
- Assume a proportional hazards model with a constant baseline hazard: $\lambda_i(t; Z_i, \theta) = \lambda \exp(\beta' Z_i)$, $i = 1, \ldots, n$
- The likelihood expression $L_i(\lambda, \beta)$, based on one individual's observation (t_i, d_i) is

$$
\begin{aligned}
L_i(\lambda, \beta) &= (\lambda e^{\beta' Z_i})^{d_i} \exp(-\lambda e^{\beta' Z_i} t_i) \\
&\propto \frac{(\lambda e^{\beta' Z_i} t_i)^{d_i} \exp(-\lambda e^{\beta' Z_i} t_i)}{d_i!} \\
&= \text{Poisson}(d_i; \lambda \exp(\beta' Z_i) t_i) \qquad (1)
\end{aligned}
$$

- The likelihood contribution is thus equivalent of modelling the event indicator D_i as a Poisson variable with expectation $\lambda \exp(\beta' Z_i) t_i$

Poissson vs. exponential regression (3)

- Poisson regression is usually treated as a generalised linear model, with the canonical log link applied to the expected value of the outcome D_i (the event indicator)
- The expected value is $\mathrm{E}(D_i) = \lambda \exp(\beta' Z_i)\mathrm{E}(\tilde{T}_i)$, where $\tilde{T}_i$ is the random time for individual i until failure **or** censoring
- Using the observed value of $\tilde{T}_i$ (which is t_i, the person-time, in our notation), we obtain Poisson model (1):

$$\log(\mathrm{E}(D_i)) \quad = \log\left[\lambda \exp(\beta' Z_i) t_i\right] = \log(\lambda) + \beta' Z_i + \overbrace{\log(t_i)}^{\textit{offset}}$$

- **N.B.** We have denoted the realised value of $\tilde{T}_i$ as t_i, the time of failure *or* censoring, which ever occurs first

Late entry

- Under the semiparametric Cox proportional hazards model, we found that individuals may be allowed to enter the risk set at a later time than 0 (late entry)
- It is next shown how such late entry can be accomodated in *parametric* survival models
- Denote the survival data as (e_i, t_i, d_i), where e_i is the entry time, t_i is the observed time of failure or censoring and d_i is the event indicator
- The likelihood contribution from an individual entering the risk set at time e_i should be conditioned on having avoided failure up to that time
  - Otherwise one would easily overestimate survival for times $< e_i$
  - In addition, we have to assume that those entering the risk set at any time $t > 0$ are statistically similar to those already in the risk set at the entry time(s); this is called *uninformative* left truncation

Late entry (2)

- The likelihood expression is derived conditionally on the individual entry times (for convenience the dependence of the hazard function on parameters θ is not indicated below):

$$
\begin{aligned}
L(\theta) &= \prod_{i=1}^{n} [\lambda(t_i)]^{d_i} S(t_i | T_i > e_i) = \prod_{i=1}^{n} \frac{\lambda(t_i)^{d_i} S(t_i)}{S(e_i)} \\
&= \prod_{i=1}^{n} \lambda(t_i)^{d_i} \frac{e^{-\int_0^{t_i} \lambda(u)du}}{e^{-\int_0^{e_i} \lambda(u)du}} = \prod_{i=1}^{n} \lambda(t_i)^{d_i} e^{-\int_{e_i}^{t_i} \lambda(u)du}
\end{aligned}
$$

- Consider individuals(s) with covariates Z_i entering the risk set at time e_i; because individuals with the same covariates but with failure time $< e_i$ would be underrepresented in the sample, situations with late entry are also referred to as left truncation ("too short" life times are underrepresented in the sample"): need conditioning to model correctly
- When the included individuals are similar to those in the risk set with respect to their future, we talk about *uninformative* left truncation

Patterns of incomplete data

- Right censoring: the failure time is known to occur after the censoring time but the exact time of failure remains unknown
- Left censoring: the failure time is known to have occurred *before* the observed time but the exact time of failure is not known
- Left trunction: the sample is not representative of failure times that have occurred before the entry time(s)
- Right truncation: the sample does not incude any information about failure times that are too long

- Right and left censoring are taken care of by likelihood expressions of the form $\mathrm{P}(T > t) = S(t)$ and $\mathrm{P}(T \leq t) = F(t)$, respectively
- Left and right truncation are taken care of by appropriate conditioning on the unit being included in the sample
- Uninformative right censoring and left truncation lead to simple analysis through control of the risk set

Estimation under censoring

- The following additional/extra material is here to show how censoring complicates the formal treatment of likelihood inference even in the simplest model
- Assume that we have i.i.d. observations from the $\mathrm{Exp}(\lambda)$ model; censoring occurs at deterministic times c_i, $i = 1, \dots, n$
- $\mathrm{E}(D_i) = 1 - \exp(-\lambda c_i)$
- The first derivative of the log likelihood is

$$\frac{d \log(L_i(\lambda))}{d\lambda} = \frac{D_i}{\lambda} - \lambda T_i$$

- The expected Fisher information is

$$-\sum_i \mathrm{E}\left[\frac{d \log L_i(\lambda)}{d\lambda^2}\right] = \sum_i \frac{\mathrm{E}(D_i)}{\lambda^2} = \frac{\sum_i (1 - e^{-\lambda c_i})}{\lambda^2}$$

Estimation under censoring (2)

- In practice, to avoid censoring times affecting the sampling properties of the ML estimator, the expected information is replaced by the observed information, so that $\sum_i d_i$ is used instead of $\mathrm{E}(D)$
- Replacing λ with the ML estimate $\hat{\lambda}$, we then get the asymptotic variance of the maximum likelihood estimator as $\hat{\lambda}^2 / \sum_i d_i$
- This was actually how we already dealt with the asymptotic standard error in Part IIA lectures (the standard error was there derived for $\log \lambda$)
- For a more detailed treatment of the topic, see Chapter 3.4.5 in Kalbfleisch and Prentice