Biostat/Epi 537

Discussion section, Week 3: Parametric models

Jan 21, 2025

Class logistics

- Homework #1 is due Friday!
- TA hours on Wednesday, 3-4 pm and Thursday, 10-11 am
- Start early so you can ask questions!

Parametric models

- How can we estimate the things we care about with survival data?
- First approach is using *parametric models*.
- Assume that the pdf of T is given by $f_T(t;\theta)$ where θ is an unknown parameter.
    - Note: θ can have multiple components/dimensions, e.g. $\theta = (\lambda, p)$
- Once we know θ, we understand the distribution of T completely.
- So, first estimate θ, then use that to estimate what we actually care about.

General workflow with parametric models

1. Specify a particular parametric form $f_T(t; \theta)$ – e.g. exponential, Weibull, gamma.
   - Choice of the model is guided by scientific knowledge.
2. Using the data, estimate θ.
   - Usually via maximum likelihood estimation (MLE) – done with software.
3. Use estimates of θ and standard error to estimate and create confidence intervals for the quantity you care about.
   - Done using the delta method.

Common survival parametric models

Distribution	Density	Survival	Hazard	Notes
Exponential(λ)	$f(t) = \lambda e^{-\lambda t}$	$S(t) = e^{-\lambda t}$	$h(t) = \lambda$	Constant hazard and memoryless property
Weibull(λ, p)	$f(t) = p\lambda(\lambda t)^{p-1}e^{-(\lambda t)^p}$	$S(t) = e^{-(\lambda t)^p}$	$h(t) = p\lambda^p t^{p-1}$	GenGamma with $\alpha = 1$
Gamma(λ, α)	$f(t) = \dfrac{\lambda^\alpha t^{\alpha-1}e^{-\lambda t}}{\Gamma(\alpha)}$	No closed form	No closed form	GenGamma with $p = 1$
GenGamma(λ, α, p)	$f(t) = \dfrac{p\lambda^{p\alpha}t^{p\alpha-1}e^{-(\lambda t)^p}}{\Gamma(\alpha)}$	No closed form	No closed form	Complicated, unintuitive

Maximum likelihood

- Data take the form (y_i, δ_i, w_i) where:
    - $y_i = \min(T_i, C_i)$ is the event time; T_i and C_i = event and censoring times
    - $\delta_i = 1$ if the observation is uncensored and 0 otherwise
    - w_i are covariates for individual i (known for everyone)
- How can we leverage information from both uncensored and censored individuals?

Maximum likelihood

You don’t need to know this to fit models, but good to know

- Impose the parametric model $f(t; \theta)$ with corresponding survival $S(t; \theta)$.
- We want to find the value of θ that is most likely to give rise to the observed data.
- If participant i is uncensored, we know $y_i = t_i$, so $f(y_i; \theta)$ captures as much information as possible about their event time.
- Otherwise, we just know that $t_i > y_i$, so the best we can do is $S(y_i; \theta) = P(T > y_i; \theta)$
- $L_n(\theta \,|\, t_1, \ldots, t_n) = \prod_{i=1}^{n} f(y_i; \theta)^{\delta_i} S(y_i; \theta)^{1-\delta_i}$
- Using calculus/software, find θ to maximize $L_n(\theta \,|\, t_1, \ldots, t_n)$ – the MLE!

Example

Exponential distribution

- When fitting survival models, you need to create a Surv object from the survival package in R.
- You can use the flexsurvreg function from the flexsurv package to fit models

```
#create surv object - necessary for fitting survival models in R
dat.surv <- Surv(time = dat$observed_times, event = dat$event, type = "right")

#fit model using flexsurv
model <- flexsurvreg(dat.surv ~ 1, dist = "exponential")
```

- You can get estimates and confidence intervals for parameters using model$res

```
> model$res
          est     L95%     U95%        se
rate 5.335709 4.946886 5.755094 0.2059828
```

Some warnings

- Some models have multiple parameterizations. Make sure you know which one R is using – if it is different from the one you are using, things can get complicated.
- When parameters are strictly positive (as is the case for the most common survival models), model$cov will return the variance/covariance estimates for the log of the parameter, not the parameter itself! This makes manually calculating confidence intervals quite tricky (more on this later).

Going beyond the parameter

The delta method (very non-rigorously)

- What we've seen so far gets us an estimate and confidence interval for λ.
- What if we care about something like an estimate/CI for $S(0.5) = P(T > 0.5)$?
- Recall: for an exponential model, $S(t; \lambda) = e^{-\lambda t}$, so $S(0.5; \lambda) = e^{-0.5\lambda}$.
- Invariance of MLE: if $\widehat{\lambda}_n$ is the MLE of λ, then $h(\widehat{\lambda}_n)$ is the MLE of $h(\lambda)$.
- So, our estimate for $S(0.5)$ is $e^{-0.5\widehat{\lambda}_n}$ where $\widehat{\lambda}_n = 5.34$ from R.

Going beyond the parameter

The delta method (very non-rigorously)

- Unfortunately, it isn't as easy to get a confidence interval.
- Need to use the *delta method.*
- If $\widehat{\theta}_n$ is the MLE for θ and $\widehat{\sigma}_n$ is the standard error, we know that $(\widehat{\theta}_n - \theta) \approx N(0, \widehat{\sigma}_n^2)$
- If we are interested instead in some function $h(\theta)$, the delta method says that $(h(\widehat{\theta}_n) - h(\theta)) \approx N(0, [h'(\widehat{\theta}_n)]^2\ \widehat{\sigma}_n^2)$
- In our case, $h(\lambda) = e^{-0.5\lambda}$.
- We could use the delta method by hand to find the SE of $h(\widehat{\theta}_n)$, and then use this to make confidence intervals.

Using software for the delta method

- Delta method requires math and can get complicated, especially with multiple dimensions.
- Luckily, there are packages in R that can do this for you!
- Using the delta method function from the msm package in R:

```r
#finding the CI using deltamethod from msm
surv_se_msm <- deltamethod(g=~exp(-0.5*x1), mean = model$res["rate", "est"],
            cov = model$res["rate", "se"]^2, ses=TRUE)
```

- This lets you get the standard error, which you can use to manually create confidence intervals.

Making life easy with fitparametric.R

- The deltamethod function can also be annoying to use
  - Still need to calculate the derivative
  - Need to pay attention to parameterizations – e.g. with Weibull
  - Need to adjust when variance-covariance matrix is given for the log of the parameter rather than the parameter, e.g. with Weibull
- The fitparametric.R script contains a number of helpful functions that can get around needing to use the delta method
  - Doesn’t cover everything, but many things that are useful for survival!

How to use fitparametric.R

What fitparametric.R can do (see R file)

- Estimate/get CI for the mean
  - What is the mean survival time?
- Estimate/get CI for survival quantiles (including median)
  - E.g. what time t is such that $P(T > t) = 0.5$ (median)?
  - E.g. what time t is such that $P(T > t) = 0.25$ (0.25 quantile)?
- Estimate/get CI for survival probabilities
  - E.g. what is $S(0.5) = P(T > 0.5)$?
- Estimate/get CI for conditional survival probabilities
  - E.g. what is $P(T > 1 \,|\, T > 0.5)$, prob. of surviving past 1 given survival past 0.5?

Recap

- Maximum likelihood estimator $\widehat{\theta_n}$ is used to estimate the parameter(s) θ of a parametric model.
    - You don't need to do this by hand or understand it deeply to fit models!
- But often, we are interested not in θ, but a function $g(\theta)$ of θ…
- Invariance of MLEs tells you that $g(\widehat{\theta}_n)$ is the MLE of $g(\theta)$ and so is a good estimator.
- Using the delta method, you can use standard error of $\widehat{\theta}_n$ to find the standard error of $g(\widehat{\theta}_n)$, and thus construct CIs for $g(\theta)$.
    - Save yourself the headache and use fitparametric.R in this class for this!