Survival/Duration/Hazard Models

Wednesday, December 6, 2023 2:00 PM

Key issue: censoring: 1) either event happens after sample ends
2) or the unit/person chooses to leave sample before the event

T = survival time / event time / failure time

C = censoring time → sample ends, or person leaves your sample

observe $Y = \min(T, C)$

$$\delta = \begin{cases} 1 & \text{if } T \le C \\ 0 & \text{if } T > C \end{cases}$$

sample (y_i, δ_i) $i = 1, \ldots, n$ units

assume: 1) only 1 event
2) C + T are independent

] both restrictive, can be relax w/ more complexity

Survival Curve or Function

$S(t) = \underline{Pr(T > t)}$ → declining in t

→ can't just calculate % of sample "alive" at each t b/c of censoring

let $d_1 < d_2 < \ldots < d_K$ be K unique event dates/times for uncensored observations

q_k = # units w/ event at d_k

r_k = # units survived until just before d_k
"at risk" units

$$Pr(T > d_k) = Pr(T > d_k \mid T > d_{k-1}) \cdot Pr(T > d_{k-1})$$

$$+ \underbrace{Pr(T > d_k \mid T \le d_{k-1})}_{=0} \cdot \cancelto{0}{Pr(T \le d_{k-1})}$$

$$= Pr(T > d_k \mid T > d_{k-1}) \cdot Pr(T > d_{k-1})$$

$$S(d_k) = \underbrace{Pr(T > d_k \mid T > d_{k-1})} \cdot S(d_{k-1})$$

$$S(d_k) = \underbrace{Pr(T > d_k \mid T > d_{k-1})}\cdot S(d_{k-1})$$

$$\hat{Pr}(T > d_k \mid T > d_{k-1}) = \frac{r_k - q_k}{r_k}$$

fraction of units "alive" at k who survive just past k

$$\hat{S}(d_k) = \prod_{j=1}^{k} \left(\frac{r_j - q_j}{r_j} \right)$$

Kaplan Meier Survival Curve

Suppose we have $(y_i, \delta_i, x_{i1}, \ldots, x_{ip})$
and want to predict survival w/ X's need to account for censoring

use hazard rate/function: likelihood of event happening in the next instant, conditional on having survived to the current instant

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t < T < t + \Delta t \mid T \geq t)}{\Delta t} \approx \frac{\partial Pr(T = t \mid T \geq t)}{\partial t}$$

$$= \frac{f(t)}{S(t)} = \frac{\partial Pr(T = t) / \partial t}{Pr(T \geq t)} = \frac{\text{unconditional}}{\text{total}}$$

likelihood of observing (y_i, δ_i).

$$L_i = \begin{cases} f(y_i) & \text{if } i \text{ not censored} \quad \text{Prob of dieing at } y_i \\ S(y_i) & \text{if } i \text{ is censored} \end{cases}$$

$$= f(y_i)^{\delta_i} \cdot S(y_i)^{1-\delta_i}$$

$$L = \prod_{i=1}^{n} f(y_i)^{\delta_i} S(y_i)^{1-\delta_i}$$

$$= \prod_{i=1}^{n} h(y_i)^{\delta_i} S(y_i) \quad \text{or} \quad h(y_i \mid \underline{x}_i)^{\delta_i} \cdot S(y_i \mid \underline{x}_i)$$

Cox Proportional Hazards Model:

Cox Proportional Hazards Model:

$$h(t \mid \underline{x}_i) = \underbrace{h_0(t)}_{\text{unspecified}} \cdot \underbrace{\exp\left(\sum_{j=1}^{p} x_{ij} \cdot \beta_j\right)}_{\text{augments baseline hazard}}$$

"relative risk" → relative to $\underline{x} = (0, 0, \dots 0)$

$$= \underbrace{\left[h_0(t) \cdot e^{\beta_0}\right]}_{\text{unspecified hazard}} \cdot e^{\overbrace{\left(\sum_{j=1}^{p} x_{ij} \beta_j\right)}^{\text{no intercept}}}$$

Suppose $\delta_i = 1$, i is uncensored $\quad y_i$ = true event time

$A_i = \cancel{h_0(y_i)} \exp(\underline{x}_i' \beta)$ hazard for i

total hazard at time y_i for all surviving units

$B_i = \sum_{i' : (y_{i'} \geq y_i)} \cancel{h_0(y_i)} \exp(\underline{x}_{i'}' \beta)$ ← all others not yet censored or not yet had event

$$\text{Prob}(i \text{ fails at time } y_i) = \frac{A_i}{B_i} = \frac{\exp(\underline{x}_i' \beta)}{\sum_{i'} (\underline{x}_{i'}' \beta)}$$

hazards cancel

Partial Likelihood $\max_{\beta} PL(\beta) = \prod_{i=1}^{n} \frac{A_i}{B_i}$ (+ shrinkage or regularization penalty)