Plan for this digital lecture

- Summary of Kaplan-Meier estimator
  - derivation of Kaplan-Meier estimator, $\widehat{A}(t)$
    - product-integral
  - estimator for variance of $\widehat{S}(t)$
  - what if tied observations
  - large sample properties of $\widehat{S}(t)$
  - estimation of median and mean survival times
- Discuss examples in ABG
  - example 3.8 (Figures 3.11 and 3.13)
  - example 3.9 (Figure 3.12)
- Confidence interval for $S(t)$
  - includes solving Problem 3.6 in ABG
- In a simple linear regression model: How to find a confidence interval for x for a given value of μ
  - related to confidence interval for the pth fractile from $\widehat{S}(t)$

Situation

- ⋆ We assume:
  - n individuals
  - each individual has the same hazard rate $\alpha(t)$ and survival function $S(t)$
  - may have truncation and/or censoring
  - $Y(t)$: number of individuals at risk just before time t

Situation

- We assume:
  - n individuals
  - each individual has the same hazard rate $\alpha(t)$ and survival function $S(t)$
  - may have truncation and/or censoring
  - $Y(t)$: number of individuals at risk just before time t

- Gives a multiplicative intensity process

$$\lambda(t) = \alpha(t)Y(t)$$

Derivation of the Kaplan-Meier estimator

- ⋆ Generalised definition of $A(t)$

$$A(t) = \int_0^t \frac{dS(u-)}{S(u-)}$$

Derivation of the Kaplan-Meier estimator

- ⋆ Generalised definition of $A(t)$

$$A(t) = \int_0^t \frac{dS(u-)}{S(u-)}$$

- ⋆ Find generalised expression for $S(t)$
  - discretise the time interval $[0, t]$
  - let the the length of the longest interval go to zero

$$S(t) = \prod_{u<t} (1 - dA(u)) = \lim_{\max(t_k - t_{k-1}) \to 0} \prod_{k=1}^{K} [1 - (A(t_k) - A(t_{k-1}))]$$

Derivation of the Kaplan-Meier estimator

- Generalised definition of $A(t)$

$$A(t) = \int_0^t \frac{dS(u-)}{S(u-)}$$

- Find generalised expression for $S(t)$
  - discretise the time interval $[0, t]$
  - let the the length of the longest interval go to zero

$$S(t) = \prod_{u<t}(1 - dA(u)) = \lim_{\max(t_k - t_{k-1}) \to 0} \prod_{k=1}^{K} [1 - (A(t_k) - A(t_{k-1}))]$$

  - If $A(t)$ is a step function, $S(t) = \sum_{u<t} \alpha_u$

$$S(t) = \prod_{u<t}(1 - \alpha_u)$$

  - If $A(t)$ is absolutely continuous

$$S(t) = \exp\{-A(t)\}$$

Derivation of the Kaplan-Meier estimator

- Generalised definition of $A(t)$

$$A(t) = \int_0^t \frac{dS(u-)}{S(u-)}$$

- Find generalised expression for $S(t)$
  - discretise the time interval $[0, t]$
  - let the the length of the longest interval go to zero

$$S(t) = \prod_{u<t}(1 - dA(u)) = \lim_{\max(t_k - t_{k-1}) \to 0} \prod_{k=1}^{K} [1 - (A(t_k) - A(t_{k-1}))]$$

  - If $A(t)$ is a step function, $S(t) = \sum_{u<t} \alpha_u$

$$S(t) = \prod_{u<t}(1 - \alpha_u)$$

  - If $A(t)$ is absolutely continuous

$$S(t) = \exp\{-A(t)\}$$

- Kaplan-Meier estimator

$$\widehat{S}(t) = \prod_{u<t}(1 - d\widehat{A}(u)) = \prod_{j: T_j \leq t} \left(1 - \frac{1}{Y(T_j)}\right)$$

Estimator for $\text{Var}[\widehat{S}(t)]$

- Using martingale theori, we found that
  - $\widehat{S}(t)$ is approximately normal
  - $\text{Var}[\widehat{S}(t)] = S^2(t)\text{Var}[\widehat{A}(t)]$

Estimator for Var$[\widehat{S}(t)]$

- ⋆ Using martingale theori, we found that
  - $\widehat{S}(t)$ is approximately normal
  - $\text{Var}[\widehat{S}(t)] = S^2(t)\text{Var}[\widehat{A}(t)]$

- ⋆ Estimator for Var$[\widehat{S}(t)]$

$$\widehat{\tau}^2(t) = \widehat{S}^2(t)\widehat{\sigma}^2(t) = \widehat{S}^2(t) \sum_{j:\, T_j \leq t} \frac{1}{Y(T_j)^2}$$

Estimator for Var$[\widehat{S}(t)]$

- Using martingale theori, we found that
  - $\widehat{S}(t)$ is approximately normal
  - $\text{Var}[\widehat{S}(t)] = S^2(t)\text{Var}[\widehat{A}(t)]$

- Estimator for Var$[\widehat{S}(t)]$

$$\widehat{\tau}^2(t) = \widehat{S}^2(t)\widehat{\sigma}^2(t) = \widehat{S}^2(t) \sum_{j: T_j \leq t} \frac{1}{Y(T_j)^2}$$

- Alternative estimator (Greenwood's formula)

$$\widehat{\tau}^2(t) = \widehat{S}^2(t) \sum_{j: T_j \leq t} \frac{1}{Y(T_j)(Y(T_j) - 1)}$$

Kaplan-Meier with tied observations

⋆ Recall

$$\widehat{S}(t) = \prod_{u<t}(1 - d\widehat{A}(u)) = \prod_{j: T_j \leq t}\left(1 - \frac{1}{Y(T_j)}\right)$$

⋆ Why tied observations:

(*i*) events happens in continuous time, but we observe ties because of rounding

(*ii*) events happens in discrete time

Kaplan-Meier with tied observations

⋆ Recall

$$\widehat{S}(t) = \prod_{u<t}(1 - d\widehat{A}(u)) = \prod_{j:T_j \leq t}\left(1 - \frac{1}{Y(T_j)}\right)$$

⋆ Why tied observations:

(*i*) events happens in continuous time, but we observe ties because of rounding

(*ii*) events happens in discrete time

⋆ Denote event times by $T_1 < T_2 < \ldots$ and multiplicities $d_1, d_2, \ldots$

⋆ Both (*i*) and (*ii*) results in

$$\widehat{S}(t) = \prod_{j:T_j \leq t}\left(1 - \frac{d_j}{Y(T_j)}\right)$$

Estimating median survival times

- ⋆ pth fractile ξ_p:

$$F(\xi_p) = p \Leftrightarrow S(\xi_p) = 1 - p$$

Estimating median survival times

- pth fractile ξ_p:

$$F(\xi_p) = p \Leftrightarrow S(\xi_p) = 1 - p$$

- Estimator for ξ_p

$$\widehat{\xi}_p = \min\{t : \widehat{S}(t) \leq 1 - p\}$$

- One can show that

$$\widehat{\mathsf{SD}}[\widehat{\xi}_p] = \frac{\widehat{\tau}(\widehat{\xi}_p)}{\widehat{f}(\widehat{\xi}_p)}$$

Estimating median survival times

- pth fractile ξ_p:

$$F(\xi_p) = p \Leftrightarrow S(\xi_p) = 1 - p$$

- Estimator for ξ_p

$$\widehat{\xi}_p = \min\{t : \widehat{S}(t) \leq 1 - p\}$$

- One can show that

$$\widehat{\text{SD}}[\widehat{\xi}_p] = \frac{\widehat{\tau}(\widehat{\xi}_p)}{\widehat{f}(\widehat{\xi}_p)}$$

- Obtain confidence interval by including in the interval all values ξ_p^0 that rejects $H_0 : \xi_p = \xi_p^0$ when tested against $H_1 : \xi_p \neq \xi_p^0$

Estimating mean survival time

- The mean is given as

$$\mu = \int_0^\infty t f(t) dt = \int_0^\infty S(t) dt$$

Estimating mean survival time

- The mean is given as

$$\mu = \int_0^\infty t f(t) dt = \int_0^\infty S(t) dt$$

- Note:
  - μ is very sensitive to the right tail behaviour
  - $\widehat{S}(t)$ typically have large variance in the right tail

Estimating mean survival time

- The mean is given as

$$\mu = \int_0^\infty t f(t) dt = \int_0^\infty S(t) dt$$

- Note:
  - μ is very sensitive to the right tail behaviour
  - $\widehat{S}(t)$ typically have large variance in the right tail
- Common to estimate the restricted mean

$$\mu_t = \int_0^t S(u) du$$

Estimating mean survival time

- The mean is given as

$$\mu = \int_0^\infty tf(t)dt = \int_0^\infty S(t)dt$$

- Note:
  - μ is very sensitive to the right tail behaviour
  - $\widehat{S}(t)$ typically have large variance in the right tail
- Common to estimate the restricted mean

$$\mu_t = \int_0^t S(u)du$$

- Estimator for μ_t

$$\widehat{\mu}_t = \int_0^t \widehat{S}(u)du$$

Estimating mean survival time

- The mean is given as

$$\mu = \int_0^\infty t f(t) dt = \int_0^\infty S(t) dt$$

- Note:
  - μ is very sensitive to the right tail behaviour
  - $\widehat{S}(t)$ typically have large variance in the right tail
- Common to estimate the restricted mean

$$\mu_t = \int_0^t S(u) du$$

- Estimator for μ_t

$$\widehat{\mu}_t = \int_0^t \widehat{S}(u) du$$

- It can be shown that an unbiased estimator for $\text{Var}[\widehat{\mu}_t]$ is

$$\widehat{\text{Var}}[\widehat{\mu}_t] = \sum_{j: T_j \leq t} \frac{(\widehat{\mu}_t - \widehat{\mu}_{T_j})^2}{Y(T_j)^2}$$