Plan for the lecture

- ⋆ Brief summary of the Nelson–Aalen estimator
  - and its properties
- ⋆ Look briefly at some examples in ABG
  - Example 3.1 (Figures 3.1, 3.2 and 3.3, pages 73, 75)
  - Example 3.2 (Figures 3.4, page 76)
- ⋆ Do two groups have different hazard rates?
  - formulated as a hypothesis test

Multiplicative intensity model

- Multiplicative intensity model

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

  - $Y(t)$: predictable process

- Frequent situation leading to a multiplicative intensity model
  - n individuals
  - each individual has the same hazard rate $\alpha(t)$
  - may have truncation and/or censoring
  - $Y(t)$: number of individuals at risk just before time t

Derivation of the Nelson–Aalen estimator

- ⋆ $N(t)$: counting process with $\lambda(t) = \alpha(t)Y(t)$
- ⋆ Used Doob-Meyer decomposition for $N(t)$ to obtain:

$$\widehat{A}(t) - A^\star(t) = \int_0^t \frac{J(s)}{Y(s)} dM(s)$$

  - $J(t) = I(Y(t) > 0)$
  - $A(t) = \int_0^t \alpha(t)dt$
  - $A^\star(t) = \int_0^t J(t)\alpha(t)dt$
  - $\widehat{A}(t) = \int_0^t \frac{J(s)}{Y(s)} dN(s)$

- ⋆ Nelson–Aalen estimator for $A(t)$:

$$\widehat{A}(t) = \int_0^t \frac{J(s)}{Y(s)} dN(s) = \sum_{j:T_j \leq t} \frac{1}{Y(T_j)}$$

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

- $\star$ Start with the martingale

$$\widehat{A}(t) - A^\star(t) = \int_0^t \frac{J(s)}{Y(s)} dM(s)$$

- $\star$ Optional variation process for $\widehat{A}(t) - A^\star(t)$

$$[\widehat{A} - A^\star](t) = \int_0^t \frac{J(s)}{Y(s)^2} dN(s) = \sum_{j: T_j \leq t} \frac{1}{Y(T_j)^2}$$

- $\star$ Since for martingales $\mathrm{Var}[M(t)] = \mathrm{E}[[M](t)]$ an unbiased estimator for $\mathrm{Var}[\widehat{A}(t) - A^\star(t)]$ is

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

Nelson–Aalen with tied observations

- Recall

$$\widehat{A}(t) = \int_0^t \frac{J(s)}{Y(s)} dN(s), \widehat{\sigma}^2(t) = \int_0^t \frac{J(s)}{Y(s)^2} dN(s)$$

- 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 $d_1, d_2, \ldots$
- Assuming (i) we get

$$\widehat{A}(t) = \sum_{j:T_j \le t} \left[\sum_{l=0}^{d_j - 1} \frac{1}{Y(T_j) - l} \right], \widehat{\sigma}^2(t) = \sum_{j:T_j \le t} \left[\sum_{l=0}^{d_j - 1} \frac{1}{(Y(T_j) - l)^2} \right]$$

- Assuming (ii) we get (we haven't discussed the reason for $\widehat{\sigma}^2(t)$)

$$\widehat{A}(t) = \sum_{j:T_j \le t} \frac{d_j}{Y(T_j)}, \widehat{\sigma}^2(t) = \sum_{j:T_j \le t} \frac{(Y(T_j) - d_j) d_j}{Y(T_j)^3}$$

Wiener process and Gaussian martingales

- ⋆ $W = \{W(t); t \geq 0\}$ is a Wiener process if
  - $W(0) = 0$
  - for any $s < t$: $W(t) - W(s) \sim N(0, t - s)$
  - independent increments
  - continuous sample paths

- ⋆ Gaussian martingale
  - let $V(t)$ be a strictly increasing continuous function with $V(0)$
  - let $W = \{W(t), t \geq 0\}$ be a Wiener process
  - let $U(t) = W(V(t))$
  - then $U(t)$ is a Gaussian martingale, i.e.
    - + $U(t)$ is a mean zero martingale
    - + $\langle U \rangle(t) = V(t)$

Rebolledo's theorem

- Theorem: Let $\widetilde{M}^{(n)}(t)$ be a sequence of mean zero martingales defined on $t \in [0, \tau]$, and assume
  - (i) $\langle \widetilde{M}^{(n)} \rangle(t) \to V(t)$ in probability when $n \to \infty$ for all $t \in [0, \tau]$
  - (ii) the sizes of the jumps of $\widetilde{M}^{(n)}$ goes to zero as $n \to \infty$

  Then $\widetilde{M}^{(n)}(t)$ converges in distribution to the mean zero Gaussian martingale $U(t) = W(V(t))$

- If
$$\widetilde{M}^{(n)}(t) = \int_0^t H^{(n)}(s) dM^{(n)}(s)$$
  where $H^{(n)}(t)$ is predictable and
$$M^{(n)}(t) = N^n(t) - \int_0^t \lambda^{(n)}(s) ds$$
  is a counting process martingale, sufficient conditions for (i) and (ii) are
  - (i) $(H^{(n)}(s))^2 \lambda^{(n)}(s) \to v(s) > 0$ when $n \to \infty$
  - (ii) $H^{(n)}(s) \to 0$ when $n \to \infty$,

  where $V(t) = \int_0^t v(s) ds$.

Large sample properties for $\widehat{A}(t)$

- Assume:
  - n individuals
  - each individual has the same hazard rate $\alpha(t)$
  - may have truncation and/or censoring
  - $Y(t)$ is number of individuals at risk just before time t
- Multiplicative intensity process: $\lambda(t) = \alpha(t)Y(t)$
- Assume also

$$\frac{Y(t)}{n} \rightarrow y(t) > 0 \text{ when } n \rightarrow \infty$$

- Then Rebolledo's theorem gives that

$$\sqrt{n}(\widehat{A}(t) - A^{\star}(t))$$

converges to a Gaussian martingale $U(t)$ with

$$\langle U \rangle(t) = \int_0^t \frac{\alpha(s)}{y(s)} ds$$

- Thus, for large n: $\widehat{A}(t) \approx N(A(t), \sigma^2(t))$