Problem 2 The Gehan-Breslow test

a) The difference is in the weighting function $L(t) = Y_1(t)Y_2(t)$, where the logrank test has $L(t) = \frac{Y_1(t)Y_2(t)}{Y_1(t)+Y_2(t)}$

Thus, while the logrank test has a weight $L(t)$ which is more or less constant, the Gehan-Breslow tests puts most weight to the first event times.

We have $\hat{A}_h(t) = \int_0^t \frac{J_h(s)}{Y_h(s)} dN_h(s)$ (Nelson-Aalen)

i.e. $d\hat{A}_h(s) = \frac{J_h(s)}{Y_h(s)} dN_h(s)$

Hence

$$Z(t_0) = \int_0^{t_0} \cancel{Y_1(s)}Y_2(s)\left[\frac{J_1(s)}{\cancel{Y_1(s)}} dN_1(s)\right]$$

$$- \int_0^{t_0} Y_1(s)\cancel{Y_2(s)}\left[\frac{J_2(s)}{\cancel{Y_2(s)}} dN_2(s)\right]$$

$$= \int_0^{t_0} Y_2(s)\,dN_1(s) - \int_0^{t_0} Y_1(s)\,dN_2(s)$$

(since $N_h(s)$ can jump only if $J_h(s) = 1$).

b) We have from general counting process theory:

$$N_h(t) = \int_0^t \lambda_h(s)\,ds + M_h(t) \leftarrow \text{Defines } M_h(t)!$$

so

$$dN_h(t) = \lambda_h(s)\,ds + dM_h(s)$$

$$= Y_h(s)\alpha_h(s)\,ds + dM_h(s)$$

$$\underset{\text{under } H_0}{=} Y_h(s)\alpha(s)\,ds + dM_h(s)$$

Hence from a),

$$Z(t_0) = \int_0^{t_0} Y_2(s) Y_1(s)\alpha(s)\,ds$$

$$+ \int_0^{t_0} Y_2(s)\,dM_1(s)$$

$$- \int_0^{t_0} Y_1(s) Y_2(s)\alpha(s)\,ds - \int_0^{t_0} Y_1(s)\,dM_2(s)$$

$$= \int_0^{t_0} Y_2(s)\,dM_1(s) - \int_0^{t_0} Y_1(s)\,dM_2(s) \quad \text{QED}$$

c) $Z(t)$ is a martingale with respect to the history $\mathcal{F}_t$ common for the two processes, since it is a difference between two martingales w.r.t. $\mathcal{F}_t$.

Now,

$$\langle Z\rangle(t) = \langle \int Y_2\, dM_1\rangle + \langle \int Y_1\, dM_2\rangle$$

$$= \int_0^t Y_2(s)^2 \lambda_1(s)\,ds + \int_0^t Y_1(s)^2 \lambda_2(s)\,ds$$

by well known formulas for $\langle \int H\, dM\rangle$

$$\underset{\text{under } H_0}{=} \int_0^t Y_2(s)^2\, Y_1(s)\alpha(s)\,ds + \int_0^t Y_1(s)^2\, Y_2(s)\alpha(s)\,ds$$

$$= \int_0^t Y_1(s)Y_2(s)\,Y_\circ(s)\alpha(s)\,ds \qquad \text{QED.}$$

d) For STK4080:
First, $\mathrm{Var}(Z(t_0)) = E\langle Z\rangle(t_0)$.

Now by Hint, $dN_\circ(s) = Y_\circ(s)\alpha(s)\,ds + dM_\circ(s)$

Thus $\langle Z\rangle(t) = \int_0^t Y_1(s)Y_2(s)\left[dN_\circ(s) - dM_\circ(s)\right]$

$$= \int_0^t Y_1(s)Y_2(s)\,dN_\circ(s) - \int_0^t Y_1(s)Y_2(s)\,dM_\circ(s) \qquad (*)$$

Taking expectation, we get

$$\mathrm{Var}(Z(t_0)) = E\left[\int_0^{t_0} Y_1(s)Y_2(s)\,dN_\circ(s)\right]$$

~~Since~~ since the last part of (*) is a mean-zero martingale.

Thus $\int_0^{t_0} Y_1(s)Y_2(s)\,dN_\circ(s) = V(t_0)$ is an unbiased estimator for $\mathrm{Var}(Z(t_0))$

So
$$Y_\bullet(s)\alpha(s)\,ds = dN_\bullet(s) - dM_\bullet(s)$$

$$\langle Z\rangle(t) = \int_0^t Y_1(s)Y_2(s)\,dN_\bullet(s) - \int_0^t Y_1(s)Y_2(s)\,dM_\bullet(s)$$

Taking expectations, we get
$$\operatorname{Var} Z(t) = E\langle Z\rangle(t) = E\int_0^t Y_1(s)Y_2(s)\,dN_\bullet(s)$$

d) For STK9080

$$Z(t)n^{-\frac{3}{2}} = \int_0^t \underbrace{Y_2(s)n^{-\frac{3}{2}}}_{H_1(s)}\,dM_1(s) - \int_0^t \underbrace{Y_1(s)n^{-\frac{3}{2}}}_{H_2(s)}\,dM_2(s)$$

To check (2.60): (following ABG p. 113)

$$H_1(t)^2\lambda_1(t) + H_2^2\lambda_2(t)$$
$$= Y_2(t)^2 n^{-3}\alpha(t)Y_1(t) + Y_1(t)^2 n^{-3}\alpha(t)Y_2(t)$$
$$\xrightarrow{P} y_2(t)^2\alpha(t)y_1(t) + y_1(t)^2\alpha(t)y_2(t) = y_1(t)y_2(t)y_\bullet(t)\alpha(t)$$

To check (2.61):
$$H_1(t) = Y_2(s)n^{-\frac{3}{2}} = \frac{Y_2(s)}{n}\cdot n^{-\frac{1}{2}} \to 0 \text{ as } n\to\infty$$

This proves the result of the subpoint

e) For STK9080

First part is same as STK4080

Last part:

$$X^2(t_0) = \frac{\left(Z(t_0)n^{-\frac{3}{2}}\right)^2}{V(t_0)n^{-3}}$$

Numerator $\xrightarrow[n\to\infty]{} N\left(0, \int_0^{t_0} y_1(s)y_2(s)y_\circ(s)ds\right)$

Denominator:

$$V(t_0)n^{-3} = \int_0^{t_0} \frac{Y_1(s)}{n} \cdot \frac{Y_2(s)}{n} \cdot \frac{dN_\circ(s)}{n}$$

$$= \int_0^{t_0} \frac{Y_1(s)}{n} \cdot \frac{Y_2(s)}{n} \cdot \frac{Y_\circ(s)\alpha(s)ds}{n}$$

$$+ \int_0^{t_0} \frac{Y_1(s)}{n} \cdot \frac{Y_2(s)}{n} \cdot \frac{dM_\circ(s)}{n} \xrightarrow{?} 0$$

$$\xrightarrow{P} \int_0^{t_0} y_1(s)y_2(s)y_\circ(s)\alpha(s)ds$$