Two sample comparison

- $N_1(t)$와 $N_2(t)$는 2개의 counting process 라고 하자.
  두 process의 intensity는 다음과 같이 주어진다.

$$\lambda_1(t) = \alpha_1(t)\, Y_1(t)$$
$$\lambda_2(t) = \alpha_2(t)\, Y_2(t)$$

- 가설 $H_0: \alpha_1(t) = \alpha_2(t)$ for all $t \in [0, t_0]$

  H_0 하에서 위험함수는 같으므로 다음과 같이 쓸 수 있다.

$$\alpha(t) = \alpha_1(t) = \alpha_2(t) \quad \text{under } H_0$$

- 누적 위험 함수의 추정량 $A_h(t) = \int_0^t \alpha_h(s)\, ds$, $h=1,2$

$$\hat{A}_h(t) = \int_0^t \frac{J_h(s)}{Y_h(s)}\, dN_h(s)$$

  여기서 $J_h(t) = I(Y_h(t) > 0)$

- Weight process $L(t)$

  $L(t)$는 음수가 아닌 "predictable" weight process 라고 정의하고 $J_1(t)J_2(t) = 0$ 이면 0이 된다.

$$J_1(t)J_2(t) = 0 \iff J_1(t) = 0 \text{ or } J_2(t) = 0$$
$$\iff Y_1(t) = 0 \text{ or } Y_2(t) = 0$$

$\iff$ 두 그룹 중 어느 하나의 위험 집단이 없는 경우

o 검정 통계량 $Z_1(t_0)$ 와 그 분산의 유도

$$Z_1(t_0) = \int_0^{t_0} L(s)\left[d\hat{A}_1(s) - d\hat{A}_2(s)\right]$$

$\Rightarrow$ 구간 $[0, t_0]$에서
가중치 $L(t)$가 적용된 누적 위험함수 차이의 합

$$Z_1(t_0) = \int_0^{t_0} L(s)\left[d\hat{A}_1(s) - d\hat{A}_2(s)\right]$$

$$= \int_0^{t_0} L(s)\left[\frac{J_1(s)}{Y_1(s)}dN_1(s) - \frac{J_2(s)}{Y_1(s)}dN_2(s)\right]$$

$\because$ $L = J_1 L = J_2 L$

$$= \int_0^{t_0} \frac{J_1(s)L(s)}{Y_1(s)}dN_1(s) - \int_0^{t_0} \frac{J_2(s)L(s)}{Y_2(s)}dN_2(s)$$

$$Z_1(t) = \boxed{\int_0^{t_0} \frac{L(s)}{Y_1(s)}dN_1(s) - \int_0^{t_0} \frac{L(s)}{Y_2(s)}dN_2(s)}$$

이때 귀무가설 H_0 하에서 $\alpha_1(t) = \alpha_2(t) = \alpha(t)$ 이고

$$dN_h(t) = \alpha(t)Y_h(t)dt + dM_h(t), \quad h = 1, 2$$

따라서

$$Z_1(t_0) = \int_0^{t_0} \frac{L(s)}{\cancel{Y_1(s)}}\left[\alpha(t)\cancel{Y_1(t)}dt + dM_1(t)\right]$$

$$- \int_0^{t_0} \frac{L(s)}{\cancel{Y_2(s)}}\left[\alpha(t)\cancel{Y_2(t)}dt + dM_2(t)\right]$$

$$= \int_0^{t_0} \frac{L(s)}{Y_1(s)}dM_1(t) - \int_0^{t_0} \frac{L(s)}{Y_2(s)}dM_2(t)$$

위의 식에서 $Z_1(t_0)$는 확률적분의 차이이고
평균이 0인 martingale 이다.

다음의 martingal 성질을 이용하라

① $\left\langle \int H\, dM \right\rangle(t) = \int_0^t H^2(s)\, \lambda(s)\, ds$

② M_1과 M_2가 orthogonal 하면, 즉

$\langle M_1, M_2 \rangle(t) = 0$ 이면

$$\left\langle \sum_{j=1}^{2} \int H_j\, dM_j \right\rangle(t) = \sum_{j=1}^{2} \int_0^t H^2(s)\, \lambda_j(s)\, ds$$

$$\langle Z_1 \rangle(t_0) = \int_0^{t_0} \left[\frac{L(s)}{Y_1(s)}\right]^2 \alpha(s)\, Y_1(s)\, ds + \int_0^{t_0} \left[\frac{L(s)}{Y_2(s)}\right]^2 \alpha(s)\, Y_2(s)\, ds$$

$$= \int_0^{t_0} \frac{L^2(s)}{Y_1(s)}\, \alpha(s)\, ds + \int_0^{t_0} \frac{L^2(s)}{Y_2(s)}\, \alpha(s)\, ds$$

$$= \int_0^{t_0} L^2(s)\, \frac{Y_1(s) + Y_2(s)}{Y_1(s)\, Y_2(s)}\, \alpha(s)\, ds$$

$$= \int_0^{t_0} L^2(s)\, \frac{Y_\bullet(s)}{Y_1(s)\, Y_2(s)}\, \alpha(s)\, ds$$

위에서 $Y_\bullet(t) = Y_1(t) + Y_2(t)$

이제 H_0 하에서 $\alpha(t)\, dt$를 추정치인

$$d\hat{A}(t) = \frac{dN_\bullet(t)}{dY_\bullet(t)} \qquad N_\bullet(t) = N_1(t) + N_2(t)$$

로 바꾸면 $\langle Z_1 \rangle(t_0)$의 추정량은 다음과 같이 얻는다.

$$V_{11}(t_0) = \widehat{\langle Z_1 \rangle (t_0)} = \int_0^{t_0} \frac{L^2(s)}{Y_1(s)\,Y_2(s)}\, dN_\bullet(s)$$

- 검정 통계량

H_0 하에서

$$U(t_0) = \frac{Z_1(t_0)}{\sqrt{V_{11}(t_0)}} \sim N(0,1)$$

또는

$$\chi^2(t_0) = \frac{Z_1^2(t)}{V_{11}(t_0)} \sim \chi_1^2$$

- Weight process의 선택

① $L(t) = Y_1(t)\,Y_2(t) / Y_\bullet(t)$

$$\Rightarrow Z_1(t_0) = \int_0^{t_0} \frac{L(s)}{Y_1(s)}\, dN_1(s) - \int_0^{t_0} \frac{L(s)}{Y_2(s)}\, dN_2(s)$$

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

$\Rightarrow$ log-rank test (Mantel-Hazel test)

• $Z_1(t_0)$의 다른형태

$N_2 = N_\bullet - N_1$ 이므로

$$Z_1(t_0) = \int_0^{t_0} \frac{L(s)}{Y_1(s)} dN_1(s) - \int_0^{t_0} \frac{L(s)}{Y_2(s)} dN_2(s)$$

$$= \int_0^{t_0} \frac{L(s)}{Y_1(s)} dN_1(s) - \int_0^{t_0} \frac{L(s)}{Y_2(s)} d(N_\bullet - N_1)$$

$$= \int_0^{t_0} L(s) \left[\frac{1}{Y_1(s)} - \frac{1}{Y_2(s)} \right] dN_1(s)$$

$$- \int_0^{t_0} \frac{L(s)}{Y_2(s)} dN_\bullet(s)$$

$$= \int_0^{t_0} L(s) \frac{Y_1(s) + Y_2(s)}{Y_1(s) Y_2(s)} dN_1(s)$$

$$- \int_0^{t_0} L(s) \frac{Y_1(s) + Y_2(s)}{Y_1(s) Y_2(s)} \cdot \frac{Y_1(s)}{Y_1(s) + Y_2(s)} dN_\bullet(s)$$

$$= \int_0^{t_0} K(s) dN_1(s) - \int_0^{t_0} K(s) \frac{Y_1(s)}{Y_\bullet(s)} dN_\bullet(s)$$

위에서 (i) $Y_\bullet(t) = Y_1(t) + Y_2(t)$

(ii) $K(t) = L(t) \cdot \dfrac{Y_\bullet(t)}{Y_1(t) Y_2(t)}$

= predictable weight Process

만약 $L(t) = \dfrac{Y_1(t) Y_2(t)}{Y_\bullet(t)}$ 이면 (log-rank test)

$K(t) = I(Y_\bullet(t) > 0)$ 이다.

따라서

$$Z_1(t_0) = \int_0^{t_0} I(Y_\bullet(s) > 0)\, dN_1(s) - \int_0^{t_0} I(Y_\bullet(s) > 0) \cdot \frac{Y_1(s)}{Y_\bullet(s)}\, dN_\bullet(s)$$

$$= \int_0^{t_0} dN_1(s) - \int_0^{t_0} \frac{Y_1(s)}{Y_\bullet(s)}\, dN_\bullet(s)$$

$$= N_1(t_0) - E_1(t_0)$$

또한 log-rank test 경우 추정량의 분산은

$$V_{11}(t_0) = \widehat{\langle Z_1 \rangle}(t_0) = \int_0^{t_0} \frac{L^2(s)}{Y_1(s)\, Y_2(s)}\, dN_\bullet(s)$$

$$= \int_0^{t_0} \frac{Y_1(s)\, Y_2(s)}{Y_\bullet^2(s)}\, dN_\bullet(s)$$