ניתוח השרדות - הצדקה להתפלגות הנורמלית האסימפטוטית

של אומד ה- KM (ע״פ Breslow + Crowley, Annals of Statistics, 1973)

מסגרת: נניח כי $T_1, \dots, T_n$ ב״ת וש״ה F (זמני האירועים)

$V_1, \dots, V_n$ ב״ת וש״ה G (זמני הצנזורה)

האנחנו צופים

$X_i = \min(T_i, V_i)$ (זמן המעקב עד אירוע או צנזורה)

$$\delta_i = \begin{cases} 1 & \text{אירוע נצפה} \\ 0 & \text{צנזורה} \end{cases}$$

נסמן

$$R(t) = \sum_{i=1}^{n} I(X_i \geq t)$$

= מס' הפרטים שעדיין עומדים במעקב בזמן t

[כאן $I(\text{תנאי}) = \begin{cases} 1 & \text{אם התנאי מתקיים} \\ 0 & \text{אם לא} \end{cases}$]

נניח כי ההתפלגות F היא רציפה, כך שהסתברות שיש שני אירועים באותה נקודת זמן שווה 0, ונתעלם מהאפשרות הזאת.

אזי ניתן לכתוב את אומד ה- KM ל- $S = 1 - F$ ואומד ה- NA ל- $\Lambda(t) = -\log S(t)$ כרשום להלן:

$$\hat{S}_{KM}(t) = \prod_{i=1}^{n} \left(1 - \frac{\delta_i}{R(X_i)}\right)^{I(X_i \leq t)}$$

$$\hat{\Lambda}_{NA}(t) = \sum_{i=1}^{n} \frac{\delta_i}{R(X_i)} I(X_i \leq t)$$

ניתן לראות כי

$$\hat{S}_{KM}(t) \doteq e^{-\hat{\Lambda}_{NA}(t)}$$

דיברנו על הקירוב הזה בכיתה.

הוכחה מדוקדקת נמצאת ב- Breslow+Crowley (Lemma 1).

כעת ניתן להראות כי

$$\hat{\Lambda}_{NA}(t) - \Lambda(t) \doteq \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\delta_i I(X_i \le t)}{r(X_i)} - c(X_i)\right]$$

באשר

$$r(u) = Pr(X_i \ge u)$$

$$c(u) = \int_0^{\min\{t,u\}} \frac{\lambda(u)}{r(u)}\, du$$

זה מוכח במאמר של Breslow + Crowley.

הטענות להגיע לזה הן מתוחכמות. יתכן שנכנס לענין זה יותר מאוחר בקורס.

בתוצאה הנ״ל $\hat{\Lambda}_{NA}(t) - \Lambda(t)$ מוצג בצורת ממוצע של המשתנים המקריים

$$W_i = \frac{\delta_i I(X_i \le t)}{r(X_i)} - c(X_i)$$

דהינו ממוצע של מ״מ ב״ת וש״ה. לכן נובע ממשפט הגבול המרכזי ש- $\hat{\Lambda}_{NA}(t) - \Lambda(t)$ מפולג בקירוב נורמלית עבור n גדול. הקירוב הנורמלי ל- $\hat{S}_{KM}(t)$ נובע מהתוצאה הזאת דרך שיטת דלתא.