"Per aspera ad astra" або нариси з аналізу виживаності та побудови критерію однорідності.

Горбунова Даніела Денисовича
4 курс бакалаврату
група "комп'ютерна статистика"

22 вересня 2020 р.

1 Вступ.

2 Початкові відомості.

2.1 Аналіз виживаності. Оцінка Каплана-Мейєра.

Нехай $\xi(t) = (\xi_1(t), \ldots, \xi_n(t))$, $t \in T$, де $t \in T = [0, b]$, b - час завершення статистичного експерименту. $\xi_1 = \xi_1(t, \omega) \in \{0, \pm 1\}, \omega \in \Omega$.

Якщо вважати, що $\xi_j(t) \neq -1$ м.н.

$$S(t) = \frac{\#\{j \in \overline{1,n} : \xi_j(t) = 0\}}{n}$$

Врахуємо кількість тих осіб, які вибули до моменту t:

$$\begin{aligned} d_t &= \#\{j \in \overline{1,n} : \xi_j(s) = 0, s \in [0,t),\ \xi_j(t) = 1\} \\ n_t &= \#\{j \in \overline{1,n} : \xi_j(s) = 0, s \in [0,t)\} \\ \hat{S}_n(t) &= \prod_{s \leqslant t} \left(1 - \frac{d_s}{n_s}\right) \end{aligned}$$

2.2 Стандартна похибка та довірчі інтервали виживаності.

Формула Грінвуда:

$$\hat{\sigma}_n^{KM}(t) = \hat{S}_n(t)\sqrt{\sum_{s\leqslant t}\frac{d_s}{n_s(n_s-d_s)}} \tag{1}$$

Для того, щоб вивести (1), доведемо дві теореми:

Теорема. Нехай послідовність випадкових величин $\{\xi_n,\ n \geqslant 1\}$ задовольняє класичну центральну граничну теорему, тобто $\exists a \in \mathbb{R} : \sqrt{n}\,(\xi_n - a) \to^W \eta \sim N(0, \sigma^2), \sigma > 0$. Тоді для довільної $g \in C^1(\mathbb{R})$ справедлива наступна збіжність:

$$\sqrt{n}\,(g(\xi_n) - g(a)) \to^W \eta_a \sim N(0, \sigma^2(g'(a))^2)$$

Доведення. За теоремою Лагранжа, існує така величина ζ_n, яка лежить між значеннями ξ_n та a:

$$g(\xi_n) = g(a) + g'(\zeta_n)(\xi_n - a) \tag{2}$$

З критерію Колмогорова ПЗВЧ, $\xi_n \to^P a = \mathbb{M}[\xi_1]$, тому $\zeta_n \to^P a$. Звідси маємо, що $g'(\zeta_n) \to^P g'(a)$. Підставимо (2) у вираз, для збіжності скористаємося теоремою Слуцького:

$$\sqrt{n}\,(g(\xi_n) - g(a)) = g'(\zeta_n)\sqrt{n}\,(\xi_n - a) \to^W g'(a)\eta \sim N(0, \sigma^2(g'(a))^2)$$

Теорема. (ЦГТ для мартингальних послідовностей)

Доведення.

Доведення формули Грінвуда. Неважко побачити, що $\forall t \in T$ оцінкою для ймовірності успіху буде відносна частота $\dfrac{d_t}{n_t} \sim Bin()$

$$\mathbb{M}[d_t/n_t] = p$$