James Robins g-formulája

KBT előadás, Lang Zsolt

Budapest, 2021. november 12.

Oksági következtetés longitudinális megfigyeléses adatokból

Megfigyeléses kohorsz vizsgálatok

Időben és betegenként változó kezelések

- Bizonyos időszakban van, máskor nincs
- A kezelések fajtái változnak
- A dózis változik
- Időben változó értékű **kovariánsok befolyásolják** a következő kezelés allokálását
- Az aktuális **kezelés hatással van az időtől függő kovariánsok későbbi értékére**

A megfigyelési időszak végén valamilyen esemény következik be

- Betegség alakul ki, gyógyulás/halálozás történik
- Folytonos skálájú jellemzőt mérünk meg

Egyszerű(bb) verzió

Feltesszük, hogy a kezelés allokálása szekvenciálisan figyelmen kívül hagyható, azaz

$$Y_{\underline{a}} \perp\!\!\!\perp A(t) | \underline{L}(t), \underline{A}(t-1).$$

A g-formula az $Y_{\underline{a}}$ **counterfactual célváltozó várható értékét** számolja ki az $Y_{\underline{A}}, \underline{A}(t), \underline{L}(t)$ **megfigyelt adatokból**.

Megjegyzés: az allokálás szekvenciálisan figyelmen kívül hagyható ugyanaz, mint hogy nincs hiányzó kovariáns. A cikk függelékében „**exchangeability**”-nek nevezték.

Állítás (g-formula)

$$E\left(Y_{\underline{a}}\right) = \int \dots \int E\left(Y_{\underline{A}} | \underline{L}(K) = \underline{l}_K, \underline{A}(K) = \underline{a}_K\right) \cdot \prod_{k=0}^{K} f\left(l_k | \underline{l}_{k-1}, \underline{a}_{k-1}\right) d\mu(l_k).$$

Bizonyítás:

$$E\left(Y_{\underline{a}}\right) = E\left(E\left(Y_{\underline{a}} | L_0, A_0\right)\right) = \int E\left(Y_{\underline{a}} | L_0 = l_0, A_0 = a_0\right) \cdot f(l_0) d\mu(l_k),$$

ui. $E\left(Y_{\underline{a}} | L_0 = l_0, A_0 = a_0\right) = E\left(Y_{\underline{a}} | L_0 = l_0\right)$.

Hasonló összefüggést írhatunk fel $E\left(Y_{\underline{a}} | L_0 = l_0, A_0 = a_0\right)$-ra, rétegezve az első időpontban. Ezt folytatva kapjuk, hogy

$$E\left(Y_{\underline{a}}\right) = \int \dots \int E\left(Y_{\underline{a}} | \underline{L}(K) = \underline{l}_K, \underline{A}(K) = \underline{a}_K\right) \cdot \prod_{k=0}^{K} f\left(l_k | \underline{l}_{k-1}, \underline{a}_{k-1}\right) d\mu(l_k).$$

Igen ám, de $E\left(Y_{\underline{a}} | \underline{L}(K) = \underline{l}_K, \underline{A}(K) = \underline{a}_K\right) = E\left(Y_{\underline{A}} | \underline{L}(K) = \underline{l}_K, \underline{A}(K) = \underline{a}_K\right)$.

Ebben a g-formula verzióban az $\underline{a} = \underline{a}_K = (a_0, a_1, \dots a_K)$ terápiasorozat előre adott, nem függ a véletlentől. Ez az ún. **statikus determinisztikus** változat.

A g-formula általánosításában az időben k-adik a_k counterfactual (tényellentétes) terápiát egy előre adott „önkényes”

$$f^{int}\left(a_k | \underline{l}_k, \underline{a}_{k-1}\right)$$

feltételes eloszlásból sorsoljuk ki, ismertnek feltételezve az adott k időpontig az összes $\underline{l}_k$ kovariánst és a $k-1$-ik időpontig az összes előző $\underline{a}_{k-1}$ terápiát.

Több változata ismert:

- **determinisztikus**, ha f^{int} elfajuló eloszlás, vagyis nincs sorsolás,
- ha nem determinisztikus, akkor **random**,
- **statikus**, ha a_k nem függ az $\underline{l}_k$ kovariánsoktól, azaz pontosabban $f^{int}\left(a_k | \underline{l}_k, \underline{a}_{k-1}\right) = f^{int}\left(a_k | \underline{a}_{k-1}\right)$,
- ha nem statikus, akkor **dinamikus**.

A g-formula általános alakja

Állítás:

$$E(Y)=\int\dots\int E\left(Y_{\underline{A}}\,|\,\underline{L}(K)=\underline{l}_K,\underline{A}(K)=\underline{a}_K\right)\cdot\prod_{k=0}^{K}f\left(l_k|\underline{l}_{k-1},\underline{a}_{k-1}\right)\cdot f^{int}\left(a_k|\underline{l}_k,\underline{a}_{k-1}\right)d\mu(l_k,a_k).$$

A bizonyítás a speciális eset kis módosításával kapható meg.