

最適治療レジメン決定のための Q 学習と A 学習

かなり前の抄読会で動的治療レジメン (DTR: Dynamic Treatment Regimes) の話題を紹介したことがあるが、その方法論に関する研究も近年の個別化医療への関心、機械学習技術の進歩などもありいくつか進展がみられる。DTR とは個々の患者特性や治療歴、臨床転帰などに基づいて治療法を連続的に決定するためのルールのことであり、ある患者に対して利用可能なすべての情報を用いてその患者に対する「最善な」アウトカムを達成することが目標となる。

DTR におけるターゲットパラメータは、 $V(d) = E\{Y^*(d)\}$, $d \in D$ であり、 $E\{Y^*(d^{opt})\} \geq E\{Y^*(d)\}$, for all $d \in D$ となる最適治療レジメン d^{opt} を求めることである。ただし、 $Y^*(d)$ は d で表現される K 回の治療方針に従った場合に観察される反事実アウトカム、 $d = \{d_1(h_1), \dots, d_k(h_k)\}$ において時点 k での治療方針 $d_k(h_k)$ はそれまでの治療・共変量歴 ($h_j: j=1, \dots, k$) で決定されるとする (D は K 回のすべての治療レジメンのクラス)。

本抄読会では、最適治療レジメン決定のために最もよく使われる 2 つの方法 (Q 学習と A 学習) について紹介する。どちらの方法も後退帰納法 (backward induction method) に基づいて最適治療レジメンを決定することは同じであるが、前者は観察データに対する同時分布 (Q 関数) の特定が必要であるのに対し、後者は時点ごとの治療効果関数 (Blip 関数) のみを特定すればよいという違いがある。Robins 流の因果推論の文脈でいえば、前者がパラメトリック g -formula におけるアウトカム回帰、後者が構造ネスト平均モデルの g -推定に対応している。

主な参考文献

- Chakraborty B, Murphy SA, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research* 2010; 19: 317-343.
- Murphy SA. Optimal dynamic treatment regimes (with discussions). *Journal of the Royal Statistical Society, Ser. B* 2003; 65: 331-366.
- Robins JM. Optimal structural nested models for optimal sequential decisions. In *Proceedings of the Second Seattle Symposium on Biostatistics*, Lin DY, Heagerty P (eds), Springer 2004; pp.189-326.
- Schulte PJ, Tsiatis AA, Laber EB, Davidian M. Q- and A-learning methods for estimating optimal dynamic treatment regimes. *Statistical Science*. 2014; 29: 640-61.
- Tsiatis AA, Davidian M, Holloway SH, and Laber EB. *Dynamic Treatment Regimes: Statistical Methods for Precision Medicine*. CRC Press. 2019.