Targeted Group Sequential Adaptive Designs

Mark van der Laan

Department of Biostatistics, University of California, Berkeley School of Public Health

Contextual multiple-bandit problem in computer science

Consider a sequence $(W_n, Y_n(0), Y_n(1))_{n\geq 1}$ of i.i.d. random variables with common probability distribution P_0^F:

- W_n, nth context (possibly high-dimensional)
- $Y_n(0)$, nth reward under action $a = 0$ (in $]0,1[$)
- $Y_n(1)$, nth reward under action $a = 1$ (in $]0,1[$)

We consider a design in which one sequentially,

- observe context W_n
- carry out randomized action $A_n \in \{0,1\}$ based on past observations and W_n
- get the corresponding reward $Y_n = Y_n(A_n)$ (other one not revealed),

resulting in an ordered sequence of dependent observations $O_n = (W_n, A_n, Y_n)$.

Goal of experiment

We want to estimate

- the optimal treatment allocation/action rule d_0:
  $d_0(W) = \arg\max_{a=0,1} E_0\{Y(a)|W\}$, which optimizes EY_d over all possible rules d.

- the mean reward under this optimal rule d_0:
  $\Psi(P_0^F) = E_0\{Y(d_0(W))\}$,

and we want

- maximally narrow valid confidence intervals (primary) "Statistical...
- minimize regret (secondary) $\frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_i(d_n))$ ... bandits"

This general contextual multiple bandit problem has enormous range of applications: e.g., on-line marketing, recommender systems, randomized clinical trials.

Targeted Group Sequential Adaptive Designs

- We refer to such an adaptive design as a particular targeted adaptive group-sequential design (van der Laan, 2008).
- In general, such designs aim at each stage to optimize a particular data driven criterion over possible treatment allocation probabilities/rules, and then use it in next stage.
- In this case, the criterion of interest is an estimator of EY_d based on past data, but, other examples are, for example, that the design aims to maximize the estimated information (i..e., minimize an estimator of the variance of efficient estimator) for a particular statistical target parameter.

Mean reward under the optimal dynamic rule

- Notation:
  - $\bar{Q}(a, W) = E_{P^F}(Y(a) \mid W)$
  - $\bar{q}(W) = \bar{Q}(1, W) - \bar{Q}(0, W)$ ("blip function")
- Optimal rule $d(\bar{Q})$:

$$d_{\bar{Q}}(W) = \arg\max_{a=0,1} \bar{Q}(a, W) = I\{\bar{q}(W) > 0\}.$$

- Mean reward under optimal dynamic rule:

$$\Psi(P^F) = E_{P^F}\left\{\bar{Q}(d(\bar{Q})(W), W)\right\}$$

Bibliography (non exhaustive!)

- Sequential designs
  - Thompson (1933), Robbins (1952)
  - specifically in the context of medical trials
    - Anscombe (1963), Colton (1963)
    - **response-adaptive designs**: Cornfield et al. (1969), Zelen (1969), many more since then
- Covariate-adjusted Response-Adaptive (CARA) designs
  - Rosenberger et al. (2001), Bandyopadhyay and Biswas (2001), Zhang et al. (2007), Zhang and Hu (2009), Shao et al (2010)... *typically* study
    - convergence of design ... in correctly specified parametric model
  - Chambaz and van der Laan (2013), Zheng, Chambaz and van der Laan (2015) concern
    - convergence of design *and* asymptotic behavior of estimator ... using (mis-specified) parametric model

Bibliography for estimation of optimal rule under iid sampling

- Zhao et al (2012), Chakraborty et al (2013), Goldberg et al (2014), Laber et al (2014), Zhao et al (2015)
- Luedtke and van der Laan (2015, 2016), based on TMLE

Sampling Strategy: Initialization

- Choose
  - a sequence $(\bar{Q}_n)_{n\geq 1}$ of estimators of $\bar{Q}_0$
  - a function $G : [-1,1] \to]0,1[$ such that, for some $t, \xi > 0$ small,
    - $|x| > \xi$ implies $G(x) = t$ if $x < 0$ and $G(x) = 1 - t$ if $x > 0$
    - $G(0) = 50\%$ and G non-decreasing
  - $G(\bar{q})$ represents a smooth approximation of the deterministic rule $W \mapsto I\{\bar{q}(W) > 0\}$ over $[-1,+1]$, bounded away from 0 and 1.
- For $i = 1, \ldots, n_0$ (initial sample size), carry out action/treatment A_i drawn from Bernoulli(50%),
  observe $O_i = (W_i, A_i, Y_i = Y_i(A_i))$

Sampling Strategy: Sequentially learn and treat

- Conditional on $O_1, \ldots, O_n$,
  1. estimate $\bar{Q}_0$ with $\bar{Q}_n$ yields $\begin{cases} \bar{q}_n(W) = \bar{Q}_n(1, W) - \bar{Q}_n(0, W) \\ d(\bar{Q}_n)(W) = I\{\bar{q}_n(W) > 0\} \end{cases}$
  2. define $g_{n+1}(W) = G(\bar{q}_n(W))$.
     Note: If $|\bar{q}_n(W)| > \xi$, then $g_{n+1}(W) \approx d(\bar{Q}_n)(W)$
- Then
  - observe W_{n+1}
  - sample A_{n+1} from Bernoulli$(g_{n+1}(W_{n+1}))$, carry out action/treatment A_{n+1}
  - get reward $Y_{n+1} = Y_{n+1}(A_{n+1})$

Estimation of outcome regression and optimal rule: Super-learning

- At each n, we can use super-learning of $\bar{Q}_0 = E_0(Y \mid A, W)$, and thereby $\bar{q}_0$ and $d_0(W) = I(\bar{q}_0(W) > 0)$.
- In Luedtke, van der Laan (2014), we propose a super-learner of d_0 based on evaluating each candidate estimator $\hat{d}$ on a cross-validated estimator of $E_0 Y_{\hat{d}}$, so that the super-learner optimizes the dissimilarity $EY_d - EY_{d_0}$. This super-learner can include candidate estimators $\hat{d}$ based on an estimator of $\bar{q}_0$, and estimators that directly estimate d_0 by reformulating it as a classification problem using standard machine learning algorithms for classification.

TMLE of Mean Reward under Optimal Rule

Given the estimator d_i based on first i observations of the optimal rule d_0, across all $i = n_0, \ldots, n$, the TMLE is defined as follows:

- Use submodel
$$\text{Logit}\,\bar{Q}_{n,\epsilon} = \text{Logit}\,\bar{Q}_n + \epsilon H(g_n, d_n),$$
where $H(g_n)(A, W) = I(A = d_n(W))/g_n(d_n(W) \mid W)$.
- Fit ϵ with *weighted* logistic regression where the i-th observation receives weight
$$\frac{g_n(A_i \mid W_i)}{g_i(A_i \mid W_i)} I(A_i = d_n(W_i)),$$
where g_i is the actual randomization used for subject i in the design.
- This results in the TMLE $\bar{Q}_n^* = \bar{Q}_{n,\epsilon_n}$ of $\bar{Q}_0 = E_0(Y \mid A, W)$, and the TMLE of $E_0 Y_{d_0}$ is thus the resulting plug-in estimator:
$$\psi_n^* = \Psi(\bar{Q}_n^*, Q_{W,n}) = \frac{1}{n} \sum_{i=1}^{n} \bar{Q}_n^*(d_n(W_i), W_i).$$

Inference for the Mean Reward under Optimal Rule

- The TMLE solves the martingale efficient score equation $0 = \sum_{i=1}^n D^*(d_n, \bar{Q}_n^*, g_i, \psi_n^*)(O_i)$, given by

$$0 = \frac{1}{n}\sum_{i=1}^n (\bar{Q}_n^*(d_n(W_i), W_i) - \psi_n^*) + \frac{I(A_i = d_n(W_i))}{g_i(A_i \mid W_i)}(Y_i - \bar{Q}_n^*(A_i, W_i)).$$

- Since $\sum_{i=1}^n D^*(d, \bar{Q}, g_i, \psi_0)(O_i)$ is a mean zero martingale sum for any $(d, \bar{Q})$, ψ_n^* can be analyzed through functional martingale central limit theorem. That is, under mild regularity conditions, asymptotically, $\psi_n^* - EY_d$ behaves as a zero-mean martingale sum

$$\frac{1}{n}\sum_{i=1}^n \left\{ \bar{Q}(d(W_i), W_i) - E_0 Y_d + \frac{I(A_i = d(W_i))}{g_i(A_i \mid W_i)}(Y_i - \bar{Q}(A_i, W_i)) \right\},$$

where (the possibly misspecified limits) $\bar{Q}$ and d are the limits of $\bar{Q}_n^*$ and d_n.

Inference: Continued

- As a consequence of the MGCLT, $\sqrt{n}(\psi_n^* - EY_d) \Rightarrow_d N(0, \sigma_0^2)$, where the asymptotic variance σ_0^2 is estimated consistently with

$$\sigma_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\bar{Q}(d(W_i), W_i) - E_0 Y_d + \frac{I(A_i = d(W_i))}{g_i(A_i \mid W_i)}(Y_i - \bar{Q}(A_i, W_i))\right\}^2$$

- Thus, an asymptotic 0.95-confidence interval for EY_d (and for the data adaptive target parameter) EY_{d_n} is given by:

$$\psi_n^* \pm 1.96\sigma_n/n^{1/2}.$$

- If d_n consistently estimates d_0, then this yields a confidence interval for $E_0 Y_{d_0}$. Either way, we obtain a valid confidence interval for the mean reward $E_0 Y_{d_n}$ under the actual rule we learned.

Concluding Remarks

- The TMLE for group-sequential adaptive designs fully preserves the integrity of randomized trials: in other words, we obtain valid inference without any reliance on model assumptions.
- The current dominating literature on this topic relies on standard MLE for (misspecified) parametric models and is thus highly problematic.
- Sequential testing, enrichment designs, and adaptive estimation of sample size, are naturally added into these targeted group-sequential adaptive designs.
- Contrary to fixed designs, these designs are able to learn and adapt along the way, serving the patients.
- The theory also applies if the choice of algorithm is set at "time" i only based on $O_1, \ldots, O_{i-1}$. However, it is important that the design stabilizes/learns as sample size increases, so modifying the adaptation strategy during design will hurt finite sample approximation of normal limit distribution.