Construction and statistical analysis of adaptive group sequential designs for randomized clinical trials

Antoine Chambaz

(MAP5, Université Paris Descartes)

joint work with Mark van der Laan

Atelier INSERM "Avancées statistiques récentes en analyse causale", 7-8 juin 2011

Introduction

Title of talk: "Construction and statistical analysis of adaptive group sequential designs for RCTs"

What do we mean? by

- group sequential design: reference to the fact that *group sequential methods* can be equally well applied on top of adaptive designs (not shown here);
- adaptive design: a RCT design that allows the investigator to dynamically modify its course through data-driven adjustment of the randomization probability based on data accrued so far.

  Prior to collection of the data, our RCT protocol specifies:
  - the parameter Ψ of scientific interest;
  - estimation framework: the confidence level α of confidence intervals for Ψ, inferential method (based on *maximum likelihood principle*).

  We wish to modify/*adapt* the *randomization probability*.

Bibliography (non exhaustive!)

Goes back to

- Thompson (Biometrika, 1933), *On the likelihood that one unknown probability exceeds another in the view of the evidence of the two samples*
- Robbins (Bull. Amer. Math. Assoc, 1952), *Some aspects of the sequential design of experiments*

and specifically in context of medical trials

- Anscombe (JASA, 1963), *Sequential medical trials*
- Colton (JASA, 1963), *A model for selecting one of two medical treatments*
- Cornfield, Halperin, Greenhouse (JASA, 1969), *An adaptive procedure for sequential clinical trials*
- Zelen (JASA, 1969), *Play-the-winner and the controlled clinical trial*

Since then, many articles devoted to *response-adapted designs*.

Our method is a Covariate-Adjusted Response Adaptive (CARA) procedure.

- Rosenberger et al. (J. Biopharm. Statist, 2001), *Covariate-adjusted response-adaptive designs for binary response*
- Bandyopadhyay and Biswas (Biometrika, 2001), *Adaptive designs for normal responses with prognostic factors*
- Zhang et al. (Ann. Statist, 2007), *Asymptotic properties of covariate-adjusted response-adaptive designs*

typically concerned convergence results *in a correctly specified parametric model*.

RCTs: Food for thoughts

What's so good about RCTs?!

- Canonical answer: "RCTs control for unknow confounders"...
- Two special features of *ideally conducted* RCTs provide more fundamental grounding:

  (A) ideal RCTs can yield causal conclusions,

  (B) ideal RCTs are self-validating.
- About (A). Three main assumptions are involved:
  1. metaphysical assumption: probabilistic dependence may have a causal explanation;
  2. all features causally relevant to the outcome other than the treatment (and its downstream effects) are identically distributed between treatment and control groups;
  3. if there is a significant difference between treatment and control groups, we can pin it down.

  If a significant difference is found, then the only possible explanation is that it is due to the treatment assignment.
- About (B). Support for assumptions 2 and 3 are built right into RCT design:
  2. blinding, random assignment, etc.
  3. reliable statistics (sample size, best possible methodology, etc.).

Yet, need to shift from "it works somewhere" to "it will work for us"!

Only medical knowledge can help fill the gap...

cf The Art of Medicine, N. Cartwright, *The Lancet, 2011*

From yesterday's talk (though you all remember)

- Question of interest:

  Is a specific treatment of interest beneficial?

- Statistical protocol (universal):
  1. *describe* as accurately as possible the observed data structure $O \sim P$ and its law $P \in \mathcal{M}$;
  2. *express* the parameter of interest under the form $\Psi(P)$;
  3. *study* the functional $\Psi : \mathcal{M} \to \mathbb{R}$;
  4. *derive* from this study *how to estimate* $\Psi(P)$;
  5. *carry out* the estimation.
- This 5-step protocol is typical of semi-parametric statistics.
- New ingredient:
  we can data-adaptively *fine-tune* the sampling scheme, hence data *are not iid*!
- In step 4, we actually follow the TMLE methodology.

Original article by van der Laan et Rubin (2006), many other since then, and forthcoming large-audience book by Rose and van der Laan (June 2011) — a chapter is devoted to TMLE for adaptive RCTs.

Balanced versus optimal iid sampling schemes

- The observed data structure is

$$O = (W, A, Y) \in \mathcal{W} \times \{0, 1\} \times \mathbb{R}, \quad O \sim P \in \mathcal{M}$$

with W *baseline covariate*, A *nature of treatment* (say drug $A = 1$ versus placebo) and Y *outcome of disease* (say time to full recovery) and $\mathcal{M}$ fully non-parametric model.

- How to *measure* the *effect* of A on Y accounting for W?
  For instance with risk difference

$$\Psi(P) = E_P\{E_P(Y|A = 1, W) - E_P(Y|A = 0, W)\}$$

such that $\Psi(P) = 0$ means *no effect* and $\Psi(P) < 0$ means *beneficial effect*.

- Causal interpretation: define *full data structure* $X = (W, Y_0, Y_1)$ (Y_a is outcome under *intervention* $A = a$) such that

(i) randomization: $A \perp X|W$ (ii) consistency: $Y = Y_A$.

Then

$$\Psi(P) = E_P\{Y_1\} - E_P\{Y_0\}.$$

Balanced versus optimal iid sampling schemes

- Likelihood under P:

$$L(P)(O) = \underbrace{P(W)}_{(1)} \times \underbrace{P(A|W)}_{(2)} \times \underbrace{P(Y|A,W)}_{(1)} = \underbrace{Q(P)(O)}_{(1)} \times \underbrace{g(P)(A|W)}_{(2)}$$

  where $g(P)$ is the *treatment mechanism* (and $Q(P)$ is the *identifiable component* of distribution of X). Q is imposed by Nature, we control g.

- Better notation: $P \equiv (Q, g)$, $\Psi(Q, g) = \Psi(Q)$.
- In a balanced RCT, $g(A|W) = g^b(A|W) = \frac{1}{2}$, the *balanced* treatment mechanism, hence *balanced sampling scheme* (Q, g^b).
- What can we gain by choosing a g that really depends on W? The Neyman allocation:

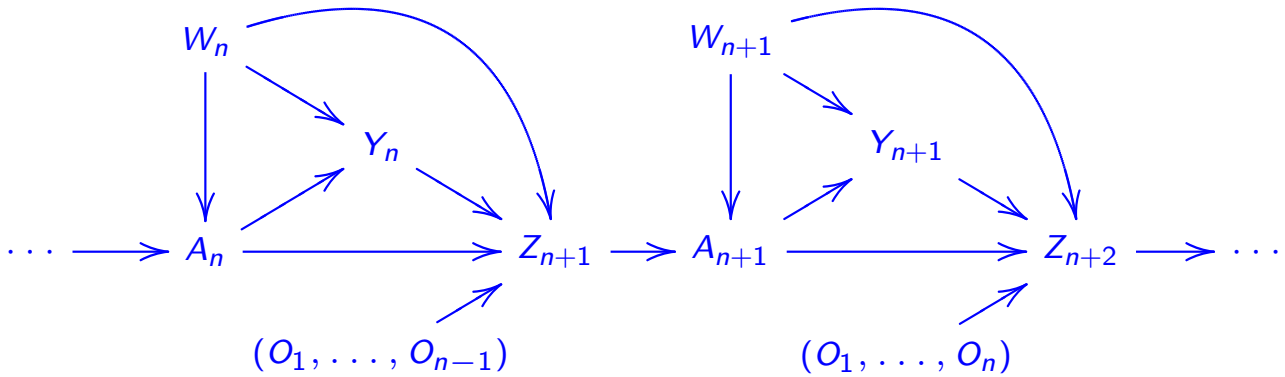
$$g^{\text{Neyman}}(Q)(A=1|W) = \frac{\sqrt{\text{Var}(Q)}(Y|A=1,W)}{\sqrt{\text{Var}(Q)}(Y|A=1,W) + \sqrt{\text{Var}(Q)}(Y|A=0,W)}$$

  minimizes the variance of all regular estimators of $\Psi(Q)$.

Sampling iid data under $(Q, g^{\text{Neyman}}(Q))$ is optimal. . . but infeasible in general!

Adaptive sampling scheme

- Choose finitely-valued $V \subset W$ and restrict the search to $\{g : g(A|W) = g(A|V)\}$.
- Rationale: at step n,



  with Z_{n+1} a finite-dimensional summary measure of the past $(O_1, \ldots, O_n)$ (depends on V).
- Only *relevant* for RCTs for which a substantial number of observations are available before all patients are randomized!
- Real-time clinical trials become feasible.

Adaptive sampling scheme: recursion-based definition

- Choose *working model* $\{Q(\theta) : \theta \in \Theta\}$ for conditional distribution of Y given (A, W). Example: $\mathcal{L}_\theta(Y|A, W) = N\left(m_\theta(A, W), \sigma_V^2(A)\right)$, m_θ identifiable.
- Yields optimal treatment mechanism $g^{\text{Neyman}}(Q(\theta)) \equiv G^{\text{Neyman}}(\theta)$.
- Draw $O_1, \ldots, O_{10} \sim^{iid} (Q_0, g^b)$; define $g_1 = \ldots = g_{10} = g^b$, $\vec{g}_{10} = (g_1, \ldots, g_{10})$, and rewrite as:

$$(O_1, \ldots, O_{10}) \sim (Q_0, \vec{g}_{10}).$$

- Say $(O_1, \ldots, O_n) \sim (Q_0, \vec{g}_n)$: by the chain rule, remarkable *factorization*

$$L(Q(\theta), \vec{g}_n)(O_1, \ldots, O_n) = \prod_{i=1}^{n} g_i(A_i|V_i) \times \prod_{i=1}^{n} Q(\theta)(O_i), \text{ hence}$$

  - weighted *maximum likelihood estimator* θ_n,
  - *adapted rd'zation probability* $g_{n+1} = \max\{\delta, \min(1 - \delta, G^{\text{Neyman}}(\theta_n))\}$ $\quad (0 < \delta << 1)$,
  - $\vec{g}_{n+1} = (\vec{g}_n, g_{n+1})$,
  - O_{n+1} drawn *conditionally on* $(O_1, \ldots, O_n)$ as

$$O_{n+1}|(O_1, \ldots, O_n) \sim (Q_0, g_{n+1}).$$

Estimation

Consistency (1/2).

Consider $(O_1, \ldots, O_n) \sim (Q_0, \vec{g}_n)$. Then the exists $\theta_0 \in \Theta$ such that, as $n \uparrow \infty$,

$$\theta_n \to \theta_0, \quad g_n = G^{\text{Neyman}}(\theta_n) \to G^{\text{Neyman}}(\theta_0), \quad \Psi(Q(\theta_n)) \to \Psi(Q(\theta_0))$$

(all convergences in probability).

- Main tool: uniform Kolmogorov strong law of large numbers for *martingales* (based on Bernstein's inequality for martingales).
- Remark: in general $G^{\text{Neyman}}(\theta_0) \neq g^{\text{Neyman}}(Q_0)$ (but can be pretty close even in the face of heavy mis-specification).

- Problem: if working model mis-specified (*i.e.*, always!) then $\Psi(Q(\theta_0)) \neq \Psi(Q_0)$ (ouch!).
- Need a *targeting step*, in order to correct the bias due to mis-specification.

Targeting step to remove bias (and increase efficiency)

- We see $Q(\theta_n)$ as an initial estimator of Q_0, that we *fluctuate* in the direction of the parameter of interest Ψ.
- Specifically, we introduce a *one-dimensional parametric model* $\{Q(\theta, \varepsilon) : (\theta, \varepsilon) \in \Theta \times \mathbb{R}\}$ such that
  - $Q(\theta, 0) = Q(\theta)$
  - the score $\frac{\partial}{\partial \varepsilon} \log Q(\theta, \varepsilon)(O)|_{\varepsilon=0}$ equals the *efficient influence curve of* Ψ *at* $Q(\theta)$ (*i.e.*, the "derivative" wrt to Q of Ψ evaluated at $Q(\theta)$).

  Example:
  - $\mathcal{L}_\theta(Y|A, W) = N\left(m_\theta(A, W), \sigma_V^2(A)\right)$,
  - define "clever covariate" $H(\theta)(A, V) = \frac{2A-1}{G^{\text{Neyman}}(\theta)(A|V)} \sigma_V^2(A)$, and introduce the *fluctuation*

$$\mathcal{L}_{(\theta,\varepsilon)}(Y|A, W) = N\left(m_\theta(A, W) + \varepsilon H(\theta)(A, V), \sigma_V^2(A)\right).$$

- Targeting step: consider the weighted maximum likelihood estimator

$$\varepsilon_n = \arg\max_{\varepsilon \in \mathbb{R}} \sum_{i=1}^{n} \log Q(\theta_n, \varepsilon)(O_i) \times g_i(A_i|V_i)^{-1}$$

hence our final estimator, the TMLE of $\Psi(Q_0)$: $\Psi(Q(\theta_n, \varepsilon_n))$.

Estimation

Consistency (2/2).

Consider $(O_1, \ldots, O_n) \sim (Q_0, \vec{g}_n)$. The TMLE $\Psi(Q(\theta_n, \varepsilon_n))$ is a consistent estimator of $\Psi(Q_0)$.

- TMLE is a *substitution estimator*: all constraints on $\Psi(Q_0)$ are necessarily satisfied.
- Keys:
  - ε_n converges in probability to some ε_0;
  - $Q(\theta_n, \varepsilon_n)$ solves the empirical "efficient influence curve equation";
  - the efficient influence curve is *double-robust*: a solution (Q, g) of the "efficient influence curve equation" at (Q_0, g) (same g's!) satisfies $\Psi(Q) = \Psi(Q_0)$ even if $Q \neq Q_0$;
  - we know the sequence $\vec{g}_n$!

Central limit theorem.

Consider $(O_1, \ldots, O_n) \sim (Q_0, \vec{g}_n)$. Then, as $n \uparrow \infty$,

$$\sqrt{n}(\Psi(Q(\theta_n, \varepsilon_n)) - \Psi(Q_0)) \Rightarrow N(0, v(\theta_0, \varepsilon_0)).$$

Furthermore, $v(\theta_0, \varepsilon_0)$ can be estimated from the data.

- Main tool: multidimensional central limit theorem for *martingales*.
- Simulations: show real gains wrt balanced design, even in the face of heavy mis-specification, for moderate and large sample sizes ($n \geq 250$).

Extensions and references

- Possible to apply *group sequential* **testing** *methods* on top of group sequential adaptive designs.
- What about applying *hypotheses selection* on top of it?
  (research in progress with Michael Rosenblum, our next speaker)

- Cartwright (2011), *The Art of Medicine*, The Lancet
- Working paper: van der Laan (2008), *The construction and analysis of adaptive group sequential designs*
- Article submitted: Chambaz, van der Laan (April 2011), *Estimation and testing in targeted goup sequential covariate-adjusted randomized clinical trials*

Merci pour votre attention !