Remarks on Estimation of and Inference on Optimal Dynamic Treatment Regimes
Discussion of Prof Butch Tsiatis' Plenary Talk

Bibhas Chakraborty
Center for Quantitative Medicine, Duke-NUS Graduate Medical School, Singapore
&
Department of Biostatistics, Columbia University

http://www.columbia.edu/∼bc2425/

SRCOS Summer Research Conference
Galveston, TX
June 2, 2014

Before I Forget ...

- It's a huge honor to serve as a discussant to Prof. Tsiatis' Plenary Talk!

- Key Collaborators:
  - Susan Murphy (Michigan)
  - Erica Moodie (McGill)
  - Eric Laber (NCSU)
  - Ying-Qi Zhao (Wisconsin-Madison)
  - Ken Cheung (Columbia)

- Supported by:
  - Start-up grant from Duke-NUS, Singapore
  - Calderone Research Prize for Junior faculty, Mailman School of Public Health, Columbia University
  - NIH grant R01 NS072127-01A1

Before I Forget ...

- It's a huge honor to serve as a discussant to Prof. Tsiatis' Plenary Talk!
- Key Collaborators:
  - Susan Murphy (Michigan)
  - Erica Moodie (McGill)
  - Eric Laber (NCSU)
  - Ying-Qi Zhao (Wisconsin-Madison)
  - Ken Cheung (Columbia)
- Supported by:
  - Start-up grant from Duke-NUS, Singapore
  - Calderone Research Prize for Junior faculty, Mailman School of Public Health, Columbia University
  - NIH grant R01 NS072127-01A1

Before I Forget ...

- It's a huge honor to serve as a discussant to Prof. Tsiatis' Plenary Talk!
- Key Collaborators:
  - Susan Murphy (Michigan)
  - Erica Moodie (McGill)
  - Eric Laber (NCSU)
  - Ying-Qi Zhao (Wisconsin-Madison)
  - Ken Cheung (Columbia)
- Supported by:
  - Start-up grant from Duke-NUS, Singapore
  - Calderone Research Prize for Junior faculty, Mailman School of Public Health, Columbia University
  - NIH grant R01 NS072127-01A1

What have we seen so far?...

- Dynamic treatment regimes (DTRs) as a vehicle to operationalize personalized medicine in a time-varying setting
  - Can be conceptualized as clinical decision support systems, a key element of the Chronic Care Model[1]
- Formalizing the key notions underlying DTRs using the potential outcomes framework and discussion of necessary assumptions (unverifiable)
- Estimation of the optimal DTR in both single stage and multistage settings
  - Value Search methods (IPWE, AIPWE) → direct methods
  - (Sequential) Regression methods (Q-learning, A-learning) → indirect methods
  - Although each method has its own issues, AIPWE came out as a more recommended approach in general

[1] Wagner EH, et al. (2001). Improving chronic illness care: Translating evidence into action. *Health Affairs*, 20: 64-78.

What have we seen so far?...

- Dynamic treatment regimes (DTRs) as a vehicle to operationalize personalized medicine in a time-varying setting
  - Can be conceptualized as clinical decision support systems, a key element of the Chronic Care Model[1]
- Formalizing the key notions underlying DTRs using the potential outcomes framework and discussion of necessary assumptions (unverifiable)
- Estimation of the optimal DTR in both single stage and multistage settings
  - Value Search methods (IPWE, AIPWE) → direct methods
  - (Sequential) Regression methods (Q-learning, A-learning) → indirect methods
  - Although each method has its own issues, AIPWE came out as a more recommended approach in general

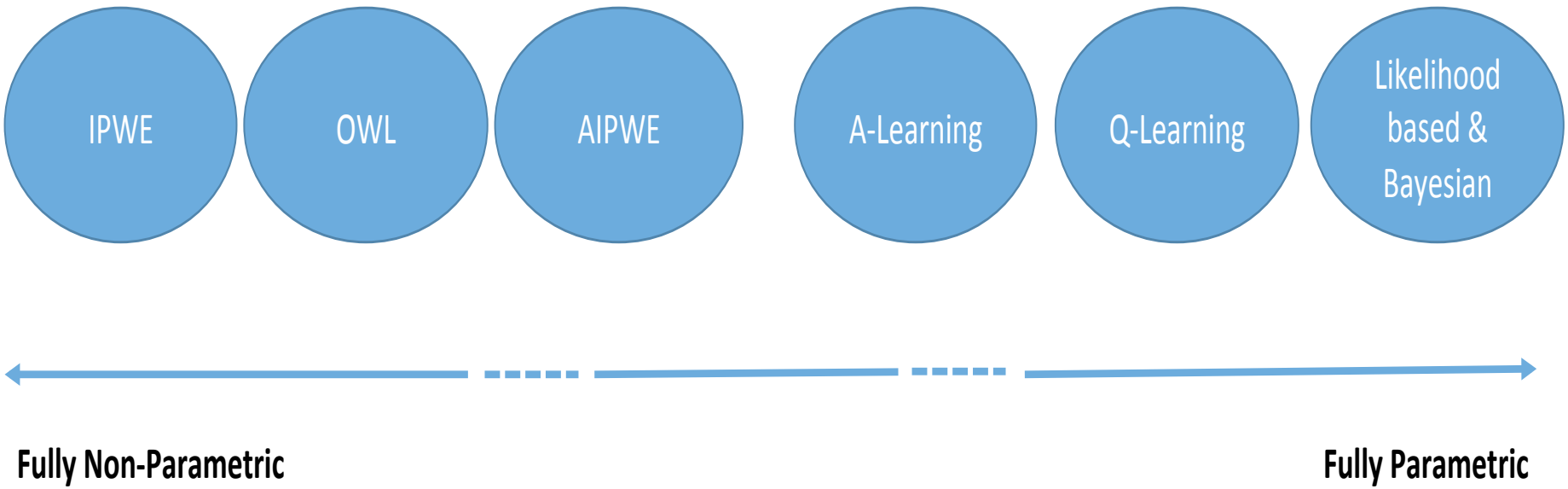
[1] Wagner EH, et al. (2001). Improving chronic illness care: Translating evidence into action. *Health Affairs*, 20: 64-78.

What have we seen so far?...

- Dynamic treatment regimes (DTRs) as a vehicle to operationalize personalized medicine in a time-varying setting
  - Can be conceptualized as clinical decision support systems, a key element of the Chronic Care Model[1]
- Formalizing the key notions underlying DTRs using the potential outcomes framework and discussion of necessary assumptions (unverifiable)
- Estimation of the optimal DTR in both single stage and multistage settings
  - Value Search methods (IPWE, AIPWE) $\rightarrow$ direct methods
  - (Sequential) Regression methods (Q-learning, A-learning) $\rightarrow$ indirect methods
  - Although each method has its own issues, AIPWE came out as a more recommended approach in general

[1] Wagner EH, et al. (2001). Improving chronic illness care: Translating evidence into action. *Health Affairs*, 20: 64-78.

Horizon of Estimation Methods



Horizon of Estimation Methods

- Q-learning and A-learning are instances of approximate dynamic programming
  - They perform well under correctly specified models, but are sensitive to model misspecification
- IPWE can reduce the "effective sample size" in estimation of Value, and can result in high variance
- AIPWE is doubly robust to model misspecification and is efficient
  - Offers a good "semi-parametric balance" between more parametric Q-learning and less parametric IPWE
  - Involves non-standard optimization and is limited to small K due to "regime consistency"

Classification Perspective in DTR

- Value search methods are closely related to classification
- Classification perspective brings some fresh air to the DTR literature
  - contrast weighted learning, related to AIPWE
  - outcome weighted learning (OWL),[2] related to IPWE
  - head-on comparison of the two?
- OWL handles the non-standard optimization problem of maximizing Value by approximating it with a surrogate convex optimization problem

[2] Zhao Y, Zeng D, Rush AJ, and Kosorok MR (2012). Estimating individualized treatment rules using outcome weighted learning. *JASA*, 107, 1106-1118.

Convex Risk Minimization in Classification

y: class label coded $-1/1$, x: set of predictors, f: the classifier

- The goal in classification is to minimize the misclassification rate involving the non-convex loss function $\mathbb{I}\{y \neq sign(f(x))\}$ or equivalently $\mathbb{I}\{y \cdot f(x) < 0\}$
- Most classification algorithms aim to minimize the expectation of a convex surrogate (convex upper bound) of the original loss function:
  - Binomial Deviance (Logistic Regression): $\log\left(1 + \exp(-2y \cdot f(x))\right)$
  - Exponential Loss (AdaBoost): $\exp(-y \cdot f(x))$
  - Hinge Loss (SVM): $(1 - y \cdot f(x))^+$
- OWL uses the hinge loss (SVM) surrogate in the context of a weighted classification problem

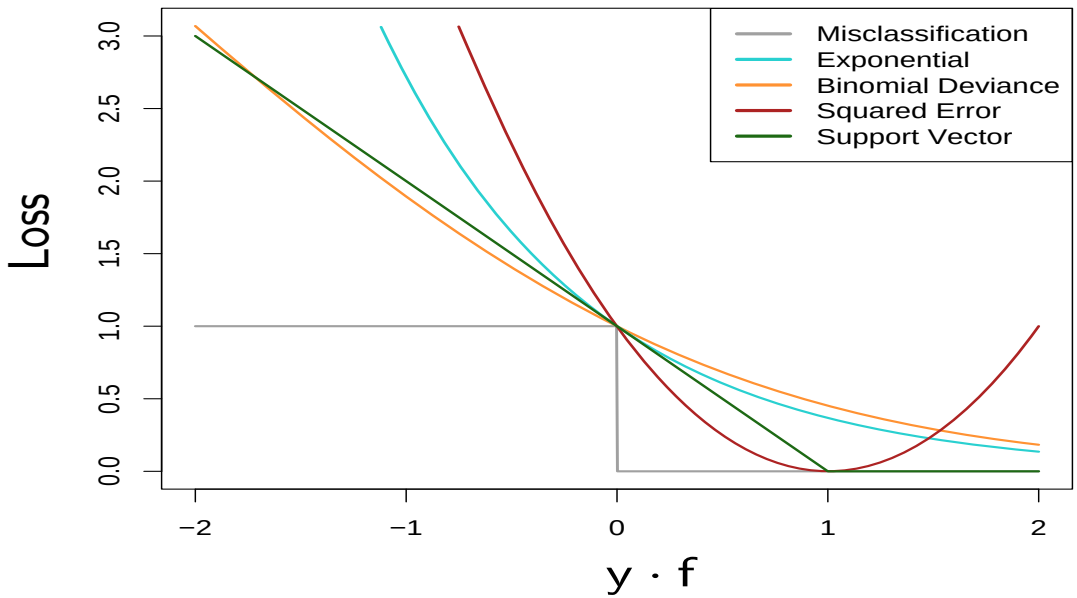


FIGURE 10.4. *Loss functions for two-class classification. The response is* $y = \pm 1$*; the prediction is* f*, with class prediction* $\text{sign}(f)$*. The losses are misclassification:* $I(\text{sign}(f) \neq y)$*; exponential:* $\exp(-yf)$*; binomial deviance:* $\log(1 + \exp(-2yf))$*; squared error:* $(y - f)^2$*; and support vector:* $(1 - yf)_+$ *(see Section 12.3). Each function has been scaled so that it passes through the point* $(0, 1)$*.*

Target of Inference?

- "Regime parameters" – the parameters that index the decision rules, e.g. parameters $\eta = (\eta_0, \eta_1)$ that index a rule:

$$d_\eta(x) = \mathbb{I}\{x_2 > \eta_0 + \eta_1 \, x_1\}$$

- If Q-learning (A-learning) is used for estimation, then regime parameters are functions of the regression parameters
  - Inference for regression parameters in Q-learning (A-learning) has been a topic of active research for last 10 years (*Robins, 2004; Moodie and Richardson, 2010; Chakraborty et al., 2010; 2013; Laber et al., 2011; Song et al., 2011*)

Target of Inference?

- "Regime parameters" – the parameters that index the decision rules, e.g. parameters $\eta = (\eta_0, \eta_1)$ that index a rule:

$$d_\eta(x) = \mathbb{I}\{x_2 > \eta_0 + \eta_1 \, x_1\}$$

- If Q-learning (A-learning) is used for estimation, then regime parameters are functions of the regression parameters
  - Inference for regression parameters in Q-learning (A-learning) has been a topic of active research for last 10 years (*Robins, 2004; Moodie and Richardson, 2010; Chakraborty et al., 2010; 2013; Laber et al., 2011; Song et al., 2011*)

Non-regularity in Inference for β_1 ($K = 2$)

$$\widehat{\tilde{Y}}_{2i} = \max\left\{Q_2(H_{2i}, 0; \hat{\beta}_2), Q_2(H_{2i}, 1; \hat{\beta}_2)\right\}$$

- Due to the non-smoothness of the "predicted Value" $\widehat{\tilde{Y}}_{2i}$ used in Q-learning, the asymptotic distribution of $\widehat{\beta}_1$ does not converge uniformly over the parameter space *(Robins, 2004)*
  - Problematic if $p > 0$, where $p \stackrel{\text{def}}{=} P[\beta_{22}^T H_2 = 0]$ (in case of linear Q-models)
  - In fact, the problem persists when $|\beta_{22}^T H_2|$ is "small" with non-zero probability (local asymptotics; Laber et al., 2011)
- Practical consequence: Both Wald type CIs and standard bootstrap CIs perform poorly
- In a K-stage setting, the same issues arise for all $\beta_k,\ k = K-1, \ldots, 1$

Target of Inference: Value of an Estimated DTR

- A more appealing target is the Value of an estimated regime $V(\widehat{d})$, irrespective of the estimation method (Q-learning, A-learning, AIPWE, ...)

- Inference on $V(\widehat{d})$ is critically important

  - Does the Value of the state-of-the-art DTR (standard of care, d_0), say $V(d_0)$, fall within the CI for the Value of the estimated DTR from the current data set, $V(\widehat{d})$?

  - For the current data set, do the estimated DTRs from two different methods (say, $\widehat{d}^{OL}$ and $\widehat{d}^{AIPWE}$) give significantly different Values?

Target of Inference: Value of an Estimated DTR

- A more appealing target is the Value of an estimated regime $V(\widehat{d})$, irrespective of the estimation method (Q-learning, A-learning, AIPWE, ...)

- Inference on $V(\widehat{d})$ is critically important

  – Does the Value of the state-of-the-art DTR (standard of care, d_0), say $V(d_0)$, fall within the CI for the Value of the estimated DTR from the current data set, $V(\widehat{d})$?

  – For the current data set, do the estimated DTRs from two different methods (say, $\widehat{d}^{QL}$ and $\widehat{d}^{AIPWE}$) give significantly different Values?

Target of Inference: Value of an Estimated DTR

- In case of $K = 2$, using the IPW idea, the Value of $\widehat{d}$ is:

$$V(\widehat{d}) = \mathbb{E}_{\hat{d}_1,\hat{d}_2} Y = \mathbb{E}\left[\frac{\mathbb{I}_{\hat{d}_1(H_1)=A_1}\mathbb{I}_{\hat{d}_2(H_2)=A_2}}{\pi_1(A_1|H_1)\pi_2(A_2|H_2)} Y\right]$$

where π_j's are the propensity scores – known by design in a SMART

- Thus $V(\widehat{d})$ is a non-smooth functional of $\widehat{d}$, or more specifically, of the estimated regime parameters $\widehat{\eta}$ (AIPWE) or $\widehat{\beta}$ (Q- or A-learning)
  - Non-smoothness increases with K
- $V(\widehat{d})$ is a data-dependent parameter – analogous to the misclassification rate of a learned classifier[3]
  - Standard inference methods do not work

[3] Laber EB and Murphy SA (2011). Adaptive confidence intervals for the test error in classification. *JASA*, 106: 904-913

Target of Inference: Value of an Estimated DTR

- In case of $K = 2$, using the IPW idea, the Value of $\widehat{d}$ is:

$$V(\widehat{d}) = \mathbb{E}_{\hat{d}_1,\hat{d}_2} Y = \mathbb{E}\left[\frac{\mathbb{I}_{\hat{d}_1(H_1)=A_1}\mathbb{I}_{\hat{d}_2(H_2)=A_2}}{\pi_1(A_1|H_1)\pi_2(A_2|H_2)} Y\right]$$

  where π_j's are the propensity scores – known by design in a SMART

- Thus $V(\widehat{d})$ is a non-smooth functional of $\widehat{d}$, or more specifically, of the estimated regime parameters $\widehat{\eta}$ (AIPWE) or $\widehat{\beta}$ (Q- or A-learning)
  - Non-smoothness increases with K

- $V(\widehat{d})$ is a data-dependent parameter – analogous to the misclassification rate of a learned classifier[3]
  - Standard inference methods do not work

[3] Laber EB and Murphy SA (2011). Adaptive confidence intervals for the test error in classification. *JASA*, 106: 904-913

Target of Inference: Value of an Estimated DTR

- In case of $K = 2$, using the IPW idea, the Value of $\widehat{d}$ is:

$$V(\widehat{d}) = \mathbb{E}_{\hat{d}_1,\hat{d}_2} Y = \mathbb{E}\left[\frac{\mathbb{I}_{\hat{d}_1(H_1)=A_1}\mathbb{I}_{\hat{d}_2(H_2)=A_2}}{\pi_1(A_1|H_1)\pi_2(A_2|H_2)} Y\right]$$

  where π_j's are the propensity scores – known by design in a SMART

- Thus $V(\widehat{d})$ is a non-smooth functional of $\widehat{d}$, or more specifically, of the estimated regime parameters $\widehat{\eta}$ (AIPWE) or $\widehat{\beta}$ (Q- or A-learning)
  - Non-smoothness increases with K

- $V(\widehat{d})$ is a data-dependent parameter – analogous to the misclassification rate of a learned classifier[3]
  - Standard inference methods do not work

[3] Laber EB and Murphy SA (2011). Adaptive confidence intervals for the test error in classification. *JASA*, 106: 904-913

m-out-of-n Bootstrap: A Feasible Solution

- m-out-of-n bootstrap is a tool for remedying bootstrap inconsistency due to non-smoothness (*Shao, 1994; Bickel et al., 1997*)
- Efron's nonparametric bootstrap with a smaller resample size, $m = o(n)$
- Choice of m has always been difficult – only a few data-driven approaches to choose m in various problems are available (*Hall et al., 1995; Lee, 1999; Cheung et al., 2005, Bickel and Sakov, 2008*)
- We developed a choice of m for the regime parameters in the context of Q-learning – adaptive to the degree of non-regularity present in the data[4]

[4] Chakraborty B, Laber EB, and Zhao Y (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics*, 69: 714 - 723.

m-out-of-n Bootstrap: A Feasible Solution

- m-out-of-n bootstrap is a tool for remedying bootstrap inconsistency due to non-smoothness (*Shao, 1994; Bickel et al., 1997*)
- Efron's nonparametric bootstrap with a smaller resample size, $m = o(n)$
- Choice of m has always been difficult – only a few data-driven approaches to choose m in various problems are available (*Hall et al., 1995; Lee, 1999; Cheung et al., 2005, Bickel and Sakov, 2008*)
- We developed a choice of m for the regime parameters in the context of Q-learning – adaptive to the degree of non-regularity present in the data[4]

[4]Chakraborty B, Laber EB, and Zhao Y (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics*, 69: 714 - 723.

Our Approach

- Key idea: Since non-regularity arises when $p > 0$, an adaptive choice of m should depend on an estimate of p
- Consider a class of resample sizes: $m = n^{\frac{1+\alpha(1-p)}{1+\alpha}}$, where $\alpha > 0$ is a tuning parameter
- Estimate p by "pre-test" of $\beta_{22}^T H_2 = 0$ for fixed H_2 over the training data set:

$$\hat{p} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\left\{ \frac{n(\hat{\beta}_{22}^T H_{2,i})^2}{H_{2,i}^T \hat{\Sigma}_2 H_{2,i}} \leq \chi_{1,1-\nu}^2 \right\}$$

- Plug in $\hat{p}$ for p in the above formula for m to get: $\hat{m} = n^{\frac{1+\alpha(1-\hat{p})}{1+\alpha}}$

Implementation

- α can be chosen in a data-driven way via double-bootstrapping (Davison and Hinkley, 1997)
- The method is robust to the choice of the pre-test level ν
- Extensive simulations support the merit of the approach over standard bootstrap
- R package qLearn: http://cran.r-project.org/web/packages/qLearn/
- Constructing one CI via double bootstrap takes about 3 minutes on a machine with dual core 2.53 GHz processor and 4GB RAM

Implementation

- α can be chosen in a data-driven way via double-bootstrapping (Davison and Hinkley, 1997)
- The method is robust to the choice of the pre-test level ν
- Extensive simulations support the merit of the approach over standard bootstrap
- R package qLearn: http://cran.r-project.org/web/packages/qLearn/
- Constructing one CI via double bootstrap takes about 3 minutes on a machine with dual core 2.53 GHz processor and 4GB RAM

Implementation

- α can be chosen in a data-driven way via double-bootstrapping (Davison and Hinkley, 1997)
- The method is robust to the choice of the pre-test level ν
- Extensive simulations support the merit of the approach over standard bootstrap
- R package qLearn: http://cran.r-project.org/web/packages/qLearn/
- Constructing one CI via double bootstrap takes about 3 minutes on a machine with dual core 2.53 GHz processor and 4GB RAM

Extension to Inference for $V(\widehat{d})$

- Key idea: In case Q-learning is used to estimate $\widehat{d}$, non-regularity for $V(\widehat{d})$ arises when $p \stackrel{\text{def}}{=} P[\min\{|\beta_{12}^T H_1|, |\beta_{22}^T H_2|\} = 0] > 0$, hence an adaptive choice of m should depend on an estimate of p

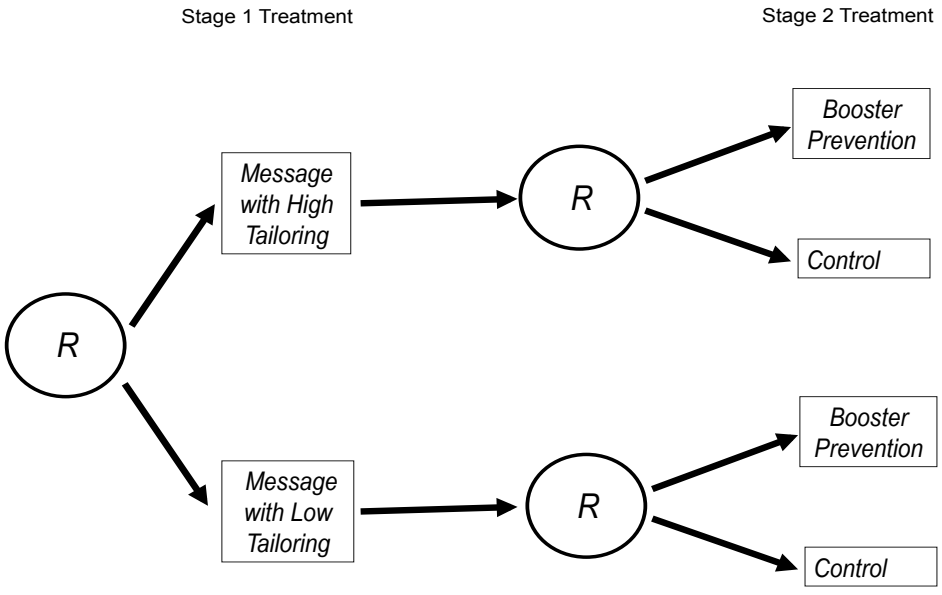
- Estimate p by "pre-test":

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}\left\{ \min\left\{ \frac{n(\hat{\beta}_{12}^T H_{1,i})^2}{H_{1,i}^T \hat{\Sigma}_1 H_{1,i}}, \frac{n(\hat{\beta}_{22}^T H_{2,i})^2}{H_{2,i}^T \hat{\Sigma}_2 H_{2,i}} \right\} \leq \chi^2_{1,1-\nu} \right\}$$

- Plug in $\hat{p}$ for p in the formula for m to get: $\hat{m} = n^{\frac{1+\alpha(1-\hat{p})}{1+\alpha}}$ and choose α via double-bootstrap

Chakraborty B, Laber EB, and Zhao Y (2014). Inference about the expected performance of a data-driven dynamic treatment regime. *Clinical Trials, in press*.

Simulating A Hypothetical SMART



A simplified version of a real smoking cessation study conducted at U-Michigan[5]

[5] Strecher VJ, et al. (2008). Web-based smoking cessation components and tailoring depth: Results of a randomized trial. *American Journal of Preventive Medicine*, 34: 373 - 381.

Simulated SMART – Data Generation

$$
\begin{aligned}
O_1 &\sim N_3(0, I); \quad O_2 = O_1 + Z, \text{ where } Z \sim N_3(0, I); \\
A_j &\sim \textit{Bernoulli}(1/2), \text{for } j = 1,\ 2; \\
Y &= O_2^T \gamma_1 + A_2 O_2^T \gamma_2 + e, \text{ where } \gamma_1 = \gamma_2 = (c, c, c)^T \text{ and } e \sim N(0, 1).
\end{aligned}
$$

The "effect size" parameter c is varied in the set $\{0.5,\ 1,\ 1.5,\ \ldots, 5\}$

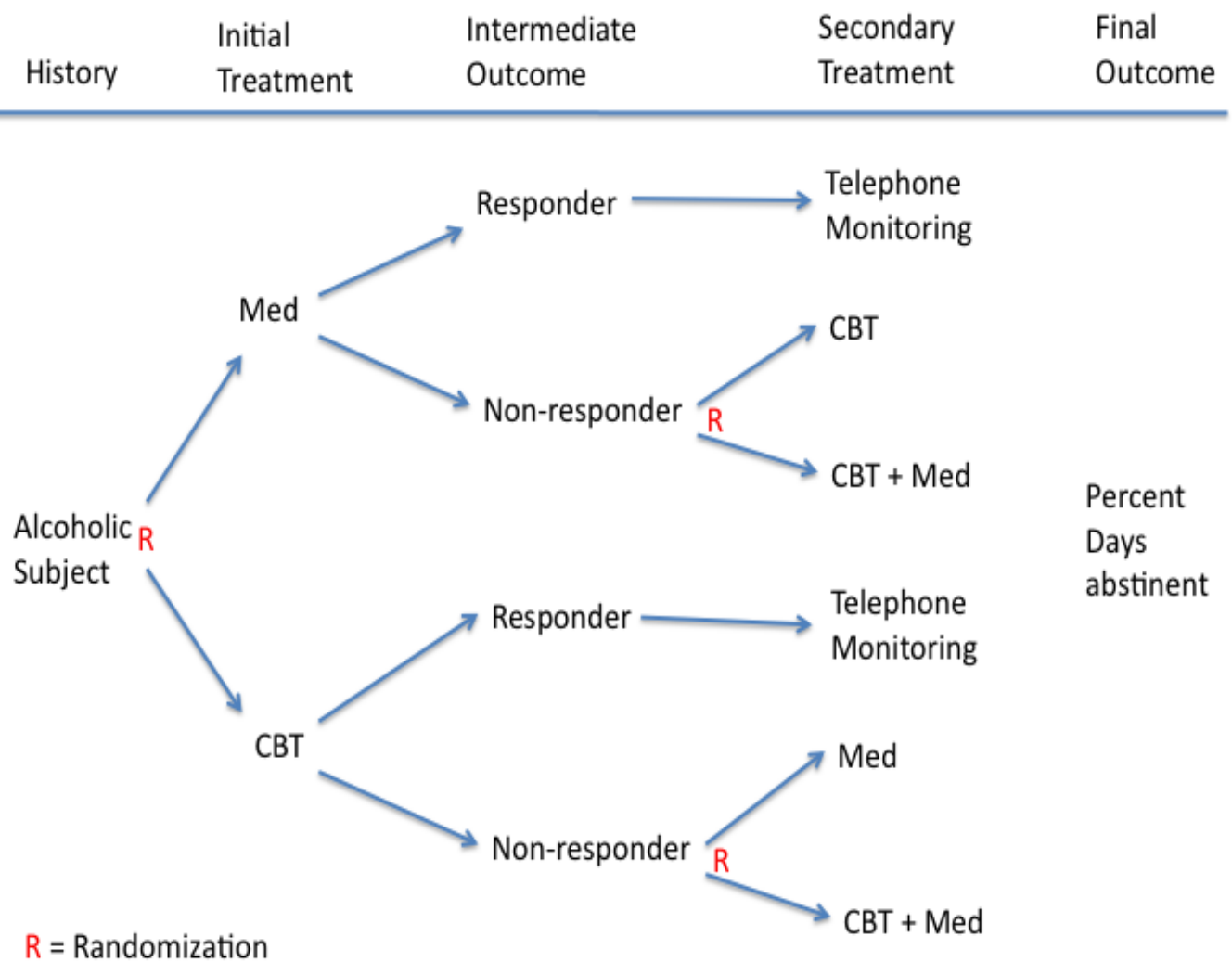
- Construct 95% nominal CI for $V(\widehat{d})$ using standard n-out-of-n and m-out-of-n bootstrap (1000 bootstrap replications)
- Monte Carlo evaluation based on 1000 simulated data sets, each of size $n = 200$

Results for $V(\widehat{d})$

$n = 200$

	Coverage		Mean Width	
c	n-out-of-n	m-out-of-n	n-out-of-n	m-out-of-n
0.5	0.866	0.934	0.976	1.197
1	0.894	0.966	1.616	1.968
1.5	0.838	0.948	2.324	2.829
2	0.882	0.952	3.036	3.771
2.5	0.894	0.956	3.791	4.642
3	0.876	0.954	4.508	5.481
3.5	0.838	0.944	5.323	6.616
4	0.858	0.944	5.898	7.188
4.5	0.854	0.942	6.686	8.138
5	0.870	0.940	7.421	9.063

Another Type of A Hypothetical SMART



Simulated SMART – Data Generation & Results

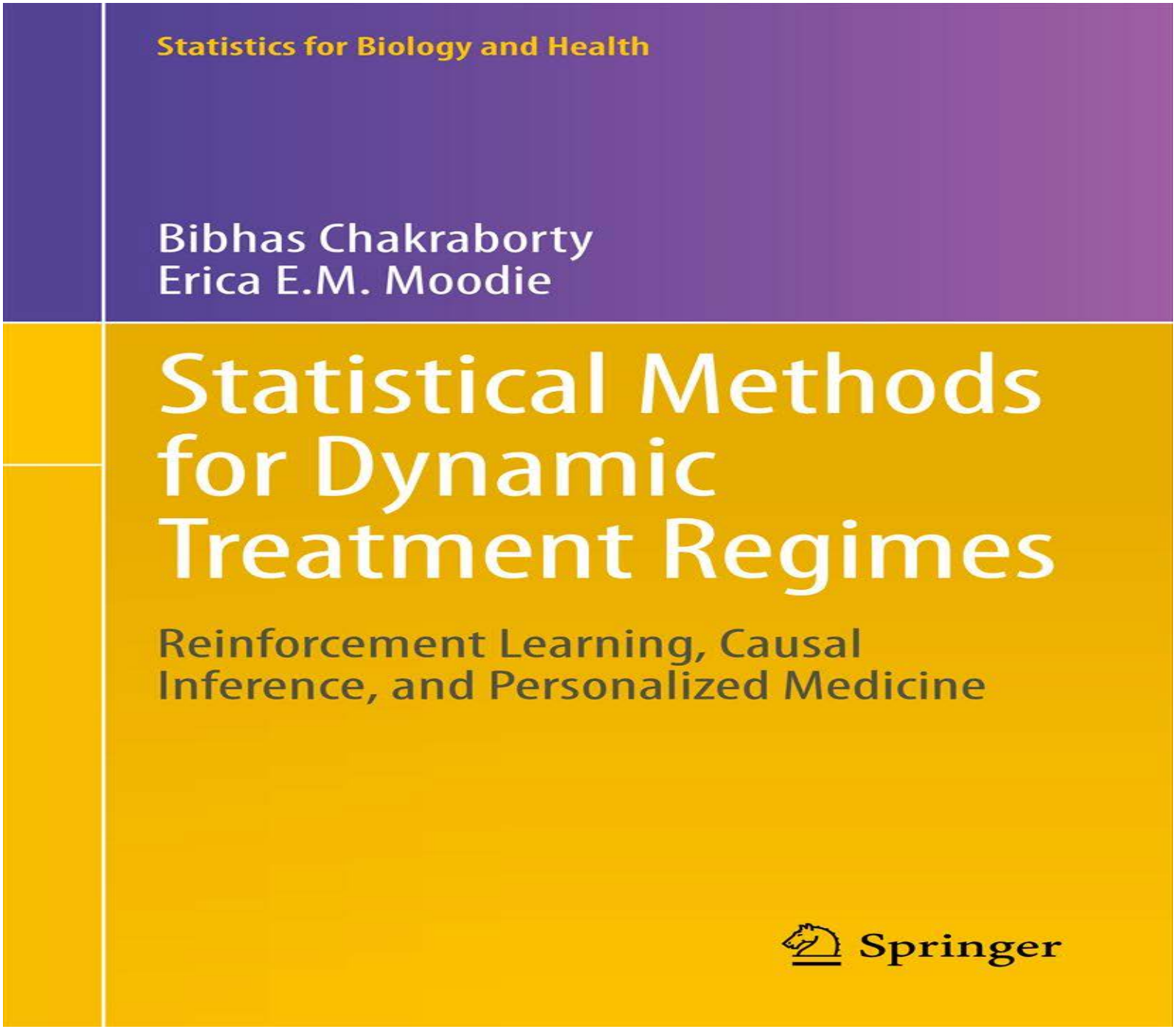
$$
\begin{aligned}
O_1 &\sim N_3(0, I); \quad O_2 = (1 + 0.05\, A_1)\, O_1 + 0.5\, Z, \text{ where } Z \sim N_3(0, I); \\
A_j &\sim \textit{Bernoulli}(1/2), \text{for } j = 1,\ 2; \\
R &= \mathbb{I}_{\{\sum_{k=1}^{3} O_{2k} \leq -1\}}\ ; \\
Y_1 &= O_1^T \gamma_1 + A_1 O_1^T \gamma_2 + e_1; \\
Y_2 &= O_2^T \gamma_1 + A_2 O_2^T \gamma_2 + e_2, \text{ where } \gamma_1 = \gamma_2 = (c, c, c)^T \text{ and } e_1,\ e_2 \overset{i.i.d.}{\sim} N(0, 1); \\
Y &= R \cdot Y_1 + (1 - R) \cdot Y_2.
\end{aligned}
$$

The effect size parameter c is set as 0.5

Similar results: 0.920 for n-out-of-n bootstrap vs. 0.942 for m-out-of-n bootstrap!

Summary & Concluding Remarks

- Inference is a difficult problem in case of Value of an estimated (data-driven) optimal DTR (since it is a non-smooth, data-dependent parameter)
  - We have proposed an adaptive m-out-of-n bootstrap scheme for constructing CIs
- Extending m-out-of-n bootstrap to more stages and more treatment choices per stage is conceptually not too problematic, but can be operationally messy
- Our data-adaptive choice of m can potentially guide such choice in other non-regular problems
  - Hodges estimator (*Hodges Jr., 1951; Beran, 1997; Samworth, 2003*)
  - Misclassification rate of a learned classifier (*Laber and Murphy, 2012*)
- Does the m-out-of-n bootstrap work for AIPWE?


Statistics for Biology and Health
Bibhas Chakraborty
Erica E.M. Moodie
Statistical Methods for Dynamic Treatment Regimes
Reinforcement Learning, Causal Inference, and Personalized Medicine
Springer

- Shoot your questions, comments, criticisms, request for slides to: bibhas.chakraborty@duke-nus.edu.sg