Precision Medicine: Lecture 04
Regression and Policy Methods

Michael R. Kosorok,
Nikki L. B. Freeman and Owen E. Leete

Department of Biostatistics
Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Fall, 2021

Outline

Regression methods

Policy learning

Set-up

- Clinical trial or observational study
- n subjects sampled from the population of interest
- Two treatment options
  - $A = 1$ observed if the patient receives the experimental treatment
  - $A = 0$ observed if the patient receives the control treatment
- X is a vector of baseline patient characteristics
- Y is the observed outcome of interest
  - Assume that larger values of Y are better

Observed Data: (Y_i, A_i, X_i), $i = 1, \ldots, n$, iid across i

The goal

The goal is to use the data, $(Y_i, A_i, X_i)_{i=1}^n$, to estimate the optimal treatment regime.

- First review what is a treatment regime and what is an optimal treatment regime
- Then see how regression can be used to estimate an optimal treatment regime
- Finally, see the conditions under which an optimal policy can be learned efficiently

Treatment regime

- A **treatment regime** is a function g that maps values of X to $\{0, 1\}$
- A patient with $X = x$ would receive treatment 1 if $g(x) = 1$ and 0 if $g(x) = 0$

Optimal treatment regime

- Optimal treatment regimes are defined in terms of potential outcomes
  - Let $Y^*(0)$ and $Y^*(1)$ denote outcomes that would be observed if treatment 0 or 1 were received by a patient, respectively
  - Assume causal consistency: $Y = Y^*(1)A + Y^*(0)(1 - A)$
  - Assume no unmeasured confounders: $\{Y^*(0), Y^*(1)\}$ independent of A conditional on X.
  - Define the potential outcome

$$Y^*(g) = Y^*(1)g(X) + Y^*(0)\{1 - g(X)\}$$

- If $\mathcal{G}$ is the class of all such regimes, the we define the optimal regime g^{opt} as

$$g^{\text{opt}} = \underset{g \in \mathcal{G}}{\text{argmax}}\, E\{Y^*(g)\}$$

Optimal treatment regime

Let $\mu(a, X) = E(Y|A = a, X)$, $a = 0, 1$. Then

$$\begin{aligned} E\{Y^*(g)\} &= E[Y^*(1)g(X) + Y^*(0)\{1 - g(X)\}] \\ &= E_X\{E[Y^*(1)g(X) + Y^*(0)\{1 - g(X)\}|X]\} \\ &= E_X[\mu(1, X)g(X) + \mu(0, X)\{1 - g(X)\}], \end{aligned}$$

and we can easily see that the optimal treatment regime is given by

$$g^{\text{opt}}(X) = I\{\mu(1, X) > \mu(0, X)\}$$

Naive Estimation

- To estimate the optimal treatment regime, an obvious approach is to posit a regression model for

$$\mu(A, X) = E(Y|A, X).$$

  - E.g., a parametric model $\mu(A, X, ; \beta)$ for finite-dimensional β
  - Use OLS or GLS to obtain an estimate $\hat{\beta}$ for β
- Assuming there is a β_0 such that $\mu(A, X) = \mu(A, X; \beta_0)$, the optimal regime is $g(X, \beta_0)$ where $g(X, \beta) = I\{\mu(1, X, \beta) > \mu(0, X, \beta)\}$
- An obvious estimator for the optimal regime is $\hat{g}^{\text{opt}}_{\text{reg}}(X) = g(X, \hat{\beta})$, and the mean outcome under the optimal regime, $E\{Y^*(g^{\text{opt}})\}$ is estimated by

$$n^{-1} \sum_{i=1}^{n} [\mu(1, X_i, \hat{\beta}) \hat{g}^{\text{opt}}_{\text{reg}}(X_i) + \mu(0, X_i, \hat{\beta})\{1 - \hat{g}^{\text{opt}}_{\text{reg}}(X_i)\}] \tag{1}$$

Class of treatment regimes

- Whether $\hat{g}^{\text{opt}}_{\text{reg}}$ is a credible estimator for g^{opt} depends on whether the model $\mu(A, X; \beta)$ is correct.
- If the conditional mean model is wrong, our estimator will not be appropriate for estimating the optimal regime
- Correct or not, the posited model $\mu(A, X; \beta)$ induces a class of treatment regimes indexed by β, say $\mathcal{G}_\beta$, with elemnts of the form $g(X, \beta)$.
- Often only a subset of X will define a regime and the class may be simplified

Class of treatment regimes

Here is an example of a model for $\mu(A, X; \beta)$

- If $\mu(A, X; \beta) = \exp(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + A(\beta_3 + \beta_4 X_1 + \beta_5 X_2))$, then elements in $\mathcal{G}_\beta$ have the form $I(\beta_3 + \beta_4 X_1 + \beta_5 X_2 > 0)$.
- Reparameterizing with $\eta_0 = -\beta_3/\beta_5$ and $\eta_1 = -\beta_4/\beta_5$, we may re-write the form of the elements of the induced class of treatment regimes as $I(X_2 > \eta_0 + \eta_1 X_1)$ or $I(X_2 < \eta_0 + \eta_1 X_1)$ depending on the sign of β_5.
- This suggests considering directly regimes of the form $g_\eta(X) = g(X, \eta)$ in a class $\mathcal{G}_\eta$.
- Outside of the regression context, considering the class indexed by η may make sense in cases such as when not all elements of X are collected in routine practice, perhaps because of cost or clinical practice.

The goal is just a missing data problem

- Thus, the goal is to find an estimator for $E[Y^*(g_\eta)]$ and estimate it directly for η to obtain an estimator $\hat{\eta}^{\text{opt}}$ for η^{opt}.
- Then $\hat{g}_\eta^{\text{opt}}(X) = g(X, \hat{\eta}^{\text{opt}})$ is an estimator for g_η^{opt}.
- For a fixed η let $C_\eta = Ag(X, \eta) + (1 - A)(1 - g(X, \eta))$.
  - When $C_\eta = 1$, $Y = Y^*(g_\eta)$
  - When $C_\eta = 0$, $Y^*(g_\eta)$ is "missing".
- We have now framed the problem as a standard missing data problem.

The goal is just a missing data problem

- Under the usual assumptions of causal consistency, SUTVA, and no unmeasured confounders, $\{Y^*(1), Y^*(0)\}$ is independend of A conditional on X
- It follows that $Y^*(g_\eta)$ is independent of C_η conditional on X. This corresponds to "missing at random" so that $\text{pr}\{C_\eta = 1 | Y^*(g_\eta), X\} = \text{pr}(C_\eta = 1 | X)$
- Letting $\pi(X) = \text{pr}(A = 1|X)$, we see that $\text{pr}(C_\eta = 1|X) = \pi_c(X;\eta) = \pi(X)g(X,\eta) + \{1 - \pi(X)\}\{1 - g(X,\eta)\}$
- In an RCT, $\pi(X)$ is known. In the observational study setting, we will need to estimate $\pi(X)$. For example, we may posit a parametric model $\pi(X;\gamma)$ (e.g. logistic regression).

IPWE

- Continuing with the missing data analogy, we can identify estimators for $E\{Y^*(g_\eta)\}$.
- For a fixed η, the inverse probability weighted estimator (IPWE) is given by

$$\begin{aligned} IPWE(\eta) &= \frac{1}{n}\sum_{i=1}^{n}\frac{C_{\eta,i}Y_i}{\pi_c(X_i;\eta,\hat{\gamma})} \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{C_{\eta,i}Y_i}{\pi(X_i;\hat{\gamma})^{A_i}\{1-\pi(X_i;\hat{\gamma})\}^{1-A_i}}. \end{aligned} \tag{2}$$

- This estimator is consistent for $E\{Y^*(g_\eta)\}$ if $\pi(X;\gamma)$, and hence $\pi_c(X;\eta,\gamma)$, is correctly specified.

AIPWE

- So far we have two estimators for $E\{Y^*(g_\eta)\}$
  - Consistency for (1) depends critically on correct specification of the conditional mean model.
  - Consistency for (2) depends critically on correct specification of propensity score model.
- An alternative estimator that provides protection against such misspecification and improves efficiency is the doubly robust augmented IPWE (AIPWE) estimator

$$AIPWE(\eta) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{C_{\eta,i}Y_i}{\pi_c(X_i;\eta,\hat{\gamma})} - \frac{C_{\eta,i}-\pi_c(X_i;\eta,\hat{\gamma})}{\pi_c(X_i;\eta,\hat{\gamma})}m(X_i;\eta,\hat{\beta})\right\} \quad (3)$$

AIPWE

- AIPWE estimator has the "doubly robust" property.
- Augmentation term increases the asymptotic efficiency when both the conditional mean model and propensity score model are correctly specified.
- The asymptotic standard error for (3) can be obtained using the usual sandwich technique.

Outline

Policy Learning

Set-up

- The goal is to learn a policy $\pi \in \Pi$ that maps a subject's features $X_i \in \mathcal{X}$ to a treatment decision: $\pi : \mathcal{X} \to \{0, 1\}$.
- Observed data is iid samples of (X_i, Y_i, W_i, Z_i) where
  - $Y_i \in \mathbb{R}$ is the outcome we want to intervene on,
  - W_i is the observed treatment assignment, and
  - Z_i is an (optional) instrument used for identifying causal effects.

  When W_i is exogenous, we take $Z_i = W_i$.
- We do not assume that the treatment propensities are known.
- For ease of exposition, we'll focus on the case when W_i is binary. See Athey and Wager (2019+) for the case of continuous W_i.

Regret and defining the best policy

The **utility** of $\pi(\cdot)$ relative to treating no one is

$$V(\pi) = E[Y_i(\pi(X_i)) - Y_i(0)]$$

where the $\{Y_i(w)\}$ correpsond to utilities we would have observed for the i-th sample had the treatment been set to $W_i = w$ and $Y_i = Y_i(W_i)$.

- This notation for potential outcomes follows Athey and Wager (2019+) and is slightly different than we had in the regression methods section.

The corresponding **policy regret** relative to the best possible policy in the class Π is

$$R(\pi) = \max\{V(\pi') : \pi' \in \Pi\} - V(\pi).$$

Policy learning goal

The goal is to learn low regret policies, i.e., to use the observed data to derive a policy $\hat{\pi} \in \Pi$ with a guarantee that $R(\hat{\pi}) = \mathcal{O}(1/\sqrt{n})$.

- To do this, we will need to make assumptions on the observational data generation distribution for identification and estimation of $V(\pi)$ (Assumptions 1 and 2).
- We will also need to control the size of Π (Assumption 3).

Assumption 1: Identification of causal effects

The primary assumption is that doubly robust scores for the average treatment effect $\theta = E[\tau(X)]$ can be constructed. Specifically:

- Let $m(x, w) = E[Y(w)|X = x] \in \mathcal{M}$ for the counterfactual response surface.
- Suppose that the induced conditional average treatment effect function $\tau_m(x)$ is linear in m.
- Suppose that we can define regret in terms of the τ-function with $V(\pi) = E[(2\pi(X) - 1)\tau_m(X)]$.
- Assume that there exists a weighting function $g(x, z)$ that identifies this τ-function

$$E[\tau_{\tilde{m}}(X_i) - g(X_i, Z_i)\tilde{m}(X_i, W_i)|X_i] = 0,$$

for any counterfactual response surface $\tilde{m}(x, w) \in \mathcal{M}$.

Identification of causal effects

With this setup, Chernozhukov et al. (2018) propose first estimating $g(\cdot)$ and $m(\cdot)$ and then considering

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}\hat{\Gamma}_i$$

$$\hat{\Gamma}_i = \tau_{\hat{m}}(X_i) + \hat{g}(X_i, Z_i)(Y_i - \hat{m}(X_i, W_i)) \tag{4}$$

to estimate θ.

Athey and Wager (2019+) use the doubly robust scores $\hat{\Gamma}_i$ as well, but rather than using them for estimating the average causal effect (policy evaluation), they use them for policy learning by plugging them into

$$\hat{\pi} = \text{argmax}\left\{\frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i) - 1)\hat{\Gamma}_i : \pi \in \Pi\right\}.$$

Assumption 2: Estimation of causal effects

With regards to estimation, only high level conditions are imposed.

- The practitioner can try different machine learning methods for each component (or combinations of methods) and use cross validation to pick the best method.
- Specific quantities are allowed to change with sample size n and dependence is noted with a subscript (e.g. $m_n(x, w)$).

Estimation of causal effects

Details for the assumptions on the estimates of m, τ_m, and g.

- Assume $E_n[m_n^2(X,W)]$, $E_n[\tau_{m_n}^2(X)]$, and $E[g_n^2(X,W)] < \infty$ for all $n = 1, 2, \ldots$.
- Assume that we have consistent estimators of the nuisance components, $\sup_{x,w}\{|\hat{m}_n(x,w) - m_n(x,w)|\} \to_p 0$, $\sup_{x}\{|\tau_{\hat{m}_n}(x) - \tau_{m_n}(x)|\} \to_p 0$, and $\sup_{x,z}\{|\hat{g}_n(x,z) - g_n(x,z)|\} \to_p 0$.
- Assume the estimators' L_2 decays as follows: for some $0 < \zeta_m, \zeta_g < 1$ with $\zeta_m + \zeta_g \geq 1$ and some $a(n) \to 0$, where X is taken to be an independent test example drawn from the same distribution as the training data:

$$E[(\hat{m}_n(X,W) - m_n(X,W))^2], E[(\tau_{\hat{m}_n}(X) - \tau_{m_n}(X))^2] \leq \frac{a(n)}{n^{\zeta_m}}$$

$$E[(\hat{g}_n(X,Z) - g_n(X,W))^2] \leq \frac{a(n)}{n^{\zeta_g}}.$$

Bounding asymptotic regret: Theorem

Given Assumption 1 and (8), and suppose that we can consistently estimate nuissance components as in Assumption 2. Suppose moreover that the irreducible noise $\epsilon_i = Y_i - m(X_i, W_i)$ is both uniformly sub-Gaussian conditionally on X_i and W_i and has second moments uniformly bounded from below, $Var[\epsilon_i | X_i = x, W_i = w] \geq s^2$, and that the treatment effect function $\tau_{m_n}(x)$ is uniformly bounded in x and n. Finally suppose that Π_n satisfies Assumption 3 with VC dimension

$$d_n = \mathcal{O}(n^\beta), \beta \leq \min \zeta_\mu, \zeta_g, \beta < 1/2.$$

Then, for any $\delta > 0$, there is a universal constant C, as well as a threshold N that depends on the constants used to define the regularity assumptions such that

$$E[R_n(\hat{\pi}_n)] \leq C \sqrt{d_n S_n^* \left(1 + \left\lfloor \log_4 \left(\frac{S_n}{S_n^*}\right) \right\rfloor\right)/n}, \text{for all } n \geq N,$$

where $R_n(\cdot)$ denotes regret for the policy based on n observations.

Assumption 3: Policy class

- To obtain regret bounds that decay as $1/\sqrt{n}$, we need to control the complexity of the class Π.
- Letting Π potentially change with n, control is achieved by assuming that Π_n is a Vapnik-Chervonenkis (VC) class with dimension that does not grow too fast with the sample size n. Specifically:
  - Assume that there is a constant $0 < \beta < 1/2$ and sequence $1 \leq d_n \leq n^{\beta}$ such that the Vapnik-Chervonenkis of Π_n is bounded by d_n for all $n = 1, 2, \ldots$.

VC Class tutorial: Motivating VC class

- Two key ideas of empirical processes theory is the "extension" of the law of large numbers and the central limit theorem.
- These ideas are used extensively in survival analysis and statistical learning theory.
- Let $X_1, \ldots, X_n$ be a random sample from a probability distribution P on a measurable space $(\mathcal{X}, \mathcal{A})$. For a measurable function $f : \mathcal{X} \mapsto \mathbb{R}$, we write $\mathbb{P}_n f$ for the expectation of f under the empirical measure and Pf for the expectation under P

$$\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^{n} f(X_i), \qquad Pf = \int f dP.$$

VC Class tutorial: Motivating VC class

- By the law of large numbers, $\mathbb{P}_n f$ converges almost surely to Pf, for every f where Pf is defined.
- The **Glivenko-Cantelli** theorems extend this result in the uniform sense in f over a class of functions.
- A class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$ is called **P-Glivenko-Cantelli** if

$$\|\mathbb{P}_n f - Pf\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |\mathbb{P}_n f - Pf| \stackrel{as*}{\to} 0.$$

VC Class tutorial: Motivating VC class

- The **empirical process** evaluated at f is defined as $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf)$.
- The multivariate central limit theorem yields the limiting distribution for any finite set of measurable functions (with finite second moments).
- The Donsker theorems extend this result in the "uniform" sense for classes of functions.
- A class $\mathcal{F}$ of measurable functions $f : \mathcal{X} \mapsto \mathbb{R}$ is called **P-Donsker** if the sequence of processes $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ converges in distribution to a tight limit process in the space $\ell^\infty(\mathcal{F})$. Then the limiting process is a P-Brownian bridge.

VC Class tutorial

- Clearly, function classes that are Glivenko-Cantelli or Donsker are very useful!
- Whether a class is Glivenko-Cantelli or Donsker depends on the "size" of the class of functions under consideration.
- Entropy calculations are strategies for measuring the size of $\mathcal{F}$. Commonly used entropy calculations are bracketing entropy and uniform entropy.
- The **Vapnik-Červonenkis classes**, or VC classes, are important examples for which good estimates on the uniform covering numbers are known.

VC Class tutorial

- A collection $\mathcal{C}$ of subsets of the sample space X **picks out** a certain set A of $\{x_1, \ldots, x_n\}$ if

$$A = C \cap \{x_1, \ldots, x_n\} \text{ for some } C \in \mathcal{C}.$$

- We say that $\mathcal{C}$ **shatters** $\{x_1, \ldots, x_n\}$ if all of the 2^n possible subsets of $\{x_1, \ldots, x_n\}$ are picked out by the sets in $\mathcal{C}$.
- The **VC-index** $V(\mathcal{C})$ of the class $\mathcal{C}$ is the smallest n for which no set of size n, $\{x_1, \ldots, x_n\} \in \mathcal{X}$, is shattered by $\mathcal{C}$.
- If $\mathcal{C}$ shatters all sets $\{x_1, \ldots, x_n\}$ for all $n \geq 1$, we set $V(\mathcal{C}) = \infty$. We say that $\mathcal{C}$ is a VC-class if $V(\mathcal{C}) < \infty$.

VC Class tutorial

Example

- Let $\mathcal{X} = \mathbb{R}$ and define the collection of sets $\mathcal{C} = \{(-\infty, c] : c \in \mathbb{R}\}$
- Consider any two point set $\{x_1, x_2\} \in \mathbb{R}$ and assume, without loss of generality, that $x_1 < x_2$.
- It is easy to verify that $\mathcal{C}$ can pick out the null set $\{\}$ and the sets $\{x_1\}$ and $\{x_1, x_2\}$ but cannot pick out $\{x_2\}$.
- Thus $V(\mathcal{C}) = 2$ and $\mathcal{C}$ is a VC-class.

VC Class tutorial

- More generally, we can define VC classes of functions.
- For a function $f : \mathcal{X} \mapsto \mathbb{R}$, the subset of $\mathcal{X} \times \mathbb{R}$ given by $\{(x, t) : t < f(X)\}$ is the subgraph of f.
- A collection $\mathcal{F}$ is a **VC class of functions** if the collection of all subgraphs of $f \in \mathcal{F}$ forms a VC class of sets in $\mathcal{X} \times \mathbb{R}$.
- We see that a collection of sets C is a VC class of sets if and only if the collection of corresponding indicator functions $\mathbf{1}_C$ is a VC class of functions. Thus it suffices to consider VC classes of functions.

Efficient policy learning: Procedure

- Divide the data into K evenly-sized folds and, for each fold $k = 1, \ldots, K$, run an estimator of choice on the other $K - 1$ data folds to estimate the function $m_n(x, w)$ and $g_n(x, z)$. Denote the estimates as $\hat{m}_n^{(-k)}(x, w)$ and $\hat{g}_n^{(-k)}(x, z)$, respectively.
- Given, these pre-computed values, choose $\hat{\pi}_n$ by maximizing a doubly robust estimate of $A(\pi) = 2V(\pi) - E[\tau(X)]$,

$$\hat{\pi} = \text{argmax}\{\hat{A}_n(\pi) : \pi \in \Pi_n\}, \tag{5}$$

$$\hat{A}_n(\pi) = \frac{1}{n}\sum_{i=1}^{n}(2\pi(X_i) - 1)\hat{\Gamma}_i, \tag{6}$$

$$\hat{\Gamma}_i = \tau_{\hat{m}_n^{(-k(i))}}(X_i) + \hat{g}_n^{(-k(i))}(X_i, Z_i)(Y_i - \hat{m}_n^{(-k(i))}(X_i, W_i)), \tag{7}$$

where $k(i) \in \{1, \ldots, K\}$ denotes the fold containing the i-th observation.

Bounding asymptotic regret

We now define some additional terms given in the main result on the asymptotic regret of policy learning using doubly robust scores. Define

$$S_n = E\left[\left(\tau_{m_n}(X_i) - g_n(X_i, Z_i)(Y_i - m_n(X_i, W_i))\right)^2\right],$$

$$S_n^* = \inf\left\{ Var[(2\pi(X_i) - 1)(\tau_{m_n}(X_i) - g_n(X_i, Z_i)(Y_i - m_n(X_i, Y_i)))] : \pi \in \Pi_n \right\}$$

where S_n bounds the second moment of the scores, and S_n^* is the asymptotic variance of (4) for estimating the policy improvement $A(\pi)$ of the best policy in Π_n.
We will also need the following assumption

$$g_n(x, z) \leq \eta \text{ for all } x, z, n \text{ and some } \eta > 0. \tag{8}$$

Bounding asymptotic regret: Synopsis

- The regret bounds on slide 24 illuminate the conditions under which we are guaranteed low regret, assuming that we can solve (5).
- The optimization task is non-convex and thus may be really hard.
- An important finding is that this type of optimization problem is numerically equivalent to a weighted classification problem, thereby connecting the policy learning problem to the tools of machine learning.
- In the next lecture, we will discuss outcome weighted learning and show how to solve (5).