

causalGLM

vanderlaanlars

September 2021

1 Introduction

2 Data-Structure and treatment-effect estimands

We will mainly consider the point-treatment data-structure $O = (W, A, Y)$ where W represents a vector of baseline covariates (i.e. possible confounders), A is a binary treatment assignment, and Y is some outcome variable. As an example, for a given observation O , W could be measurements: age, sex, a risk-score, location, income; A could take the value 1 if the individual receives the treatment and 0 if they do not receive the treatment; and Y is a binary or continuous variable that captures the effect of the treatment.

For the goal of assessing heterogeneity of the treatment effect, there are a number of popular estimands: The conditional average treatment effect (CATE):

$$CATE(w) := E[Y|A = 1, W = w] - E[Y|A = 0, W = w], \quad (1)$$

which is an additive measure of the effect of the treatment ($A = 1$) relative to no treatment ($A = 0$). The conditional odds ratio (OR) for when Y is binary:

$$OR(w) := \frac{P(Y = 1|A = 1, W = w)P(Y = 0|A = 1, W = w)}{P(Y = 1|A = 0, W = w)P(Y = 0|A = 0, W = w)} \quad (2)$$

The conditional relative risk (RR) for when Y is nonnegative (e.g. a binary or count variable):

$$RR(w) := \frac{E[Y|A = 1, W = w]}{E[Y|A = 0, W = w]}, \quad (3)$$

which is a relative measure of the effect of the treatment ($A = 1$) relative to no treatment ($A = 0$).

In some application, non-contrast measures like the conditional treatment-specific mean (TSM) may be of interest:

$$TSM_a(w) := E[Y|A = a, W = w]. \quad (4)$$

2.1 Conventional estimators using parametric generalized linear models

In order to estimate the estimands of the previous section, parametric generalized linear models are often employed (e.g. the R package *glm*). For the CATE, the following linear regression model is often used:

$$E[Y|A, W = w] = \beta_0 A + \beta_1^T W \cdot A + \beta_2^T W.$$

This model is equivalent to assuming both the nuisance linear model

$$E[Y|A = 0, W = w] = \beta_1^T W$$

and target linear model

$$CATE(w) = \beta_0 + \beta_1^T W.$$

Thus, the coefficient in front of the treatment interactions can be directly interpreted as a measure of the conditional average treatment effect when the parametric model is correct. However, we see that very strong parametric assumptions are made on the orthogonal nuisance function $w \mapsto E[Y|A = 0, W = w]$, which has little to do with the CATE.

Next, for the conditional odds ratio, the following logistic regression model is often used

$$\text{logit}\{P(Y = 1|A, W)\} = \beta_0 A + \beta_1^T W \cdot A + \beta_2^T W.$$

This model is equivalent to assuming both the nuisance logistic model

$$\text{logit}\{P(Y = 0|A = 0, W)\} = \beta_2^T W$$

and the target logistic model

$$\log OR(w) = \beta_0 + \beta_1^T W.$$

Once again, we see that the conventional logistic regression model makes strong parametric assumptions on the orthogonal nuisance parameter $P(Y = 0|A = 0, W = w)$.

Finally, for the conditional relative-risk, the poisson or log-linear regression model is a well-known approach:

$$\log\{E[Y|A, W]\} = \beta_0 A + \beta_1^T W \cdot A + \beta_2^T W.$$

This is equivalent to assuming the nuisance log-linear model

$$\log\{E[Y|A = 0, W]\} = \beta_2^T W$$

and the target log-linear model

$$\log\{RR(W)\} = \beta_0 + \beta_1^T W.$$

Besides the strong parametric assumptions on the nuisance parameter $E[Y|A = 0, W = w]$, another issue with this approach is that conventional methods only provide inference when the outcome is Poisson distributed, which is not useful if the outcome is binary. Log-link binomial generalized-linear-models are one way to overcome this limitation.

Under the parametric assumptions, standard generalize linear model software can be used to obtain estimates and inference for the coefficients in the above models for the treatment-effect estimands. However, these methods make much stronger assumptions than are needed. In particular, the parametric assumptions on $E[Y|A = 0, W]$ provides little-to-no benefit in interpretability since we are interested in the coefficients for the treatment interaction terms, and it comes at a possibly substantial cost in bias due to model misspecification. Additionally, these methods do not allow for any adaptive estimation of $E[Y|A = 0, W]$ (e.g. using the LASSO, variable selection, or machine-learning) and therefore do not perform well in both estimation and inference in high dimensions. In the next section, we consider a partial relaxation of the parametric models through so-called partially-linear generalized linear models.

3 Semiparametric generalized linear models for treatment effect estimation with causalGLMsp

In this section, we give an overview of semiparametric treatment-effect estimation in partially-linear generalized linear models which allows for adaptive estimation of nuisance parameters of the data-generating distribution that are not directly relevant for the problem at end. Semiparametric models are statistical models in which one component of the data-generating distribution is parametric and the remaining components are nonparametric.

These methods allow for:

1. Interpretable (coefficient-based) inference for user-specified parametric models for conditional treatment effect estimands.
2. Adaptive machine-learning and variable selection methods including generalized additive models, LASSO, MARS and gradient-boosting can be used to estimate nuisance parameters nonparametrically, thereby substantially relaxing assumptions for valid inference although still assuming a parametric model for the conditional estimand of interest.

3.1 Conditional average treatment effect (CATE) and partially-linear least-squares regression

The partially-linear least-squares model is of the form:

$$E[Y|A, W] = \beta_0 + \beta_1^T W \cdot A + h_0(W),$$

where $h_0(W) := E[Y|A = 0, W]$ is an unspecified nuisance function that is to be learned from the data nonparametrically. The parametric component of the model is the coefficient vector (β_0, β_1) . This model is equivalent to *only* assuming:

$$CATE(w) = \beta_0 + \beta_1^T w.$$

Thus, this semiparametric model only makes assumptions that directly relate the estimand of interest!

Estimates and inference for the coefficient vector in this semiparametric model can be obtained by applying the R function *causalGLM* with the option 'estimand = "CATE"'. We employ machine-learning for initial estimation of the relevant components of the data-generating distribution and then use targeted maximum likelihood estimation for bias-correction, thereby allowing for valid efficient inference.

3.2 Conditional odds ratio (OR) and partially-linear logistic regression

The partially-linear logistic model is of the form:

$$\text{logit}\{E[Y|A, W]\} = \beta_0 + \beta_1^T W \cdot A + h_0(W),$$

where $h_0(W) := \text{logit}\{E[Y|A = 0, W]\}$ is an unspecified nuisance function that is to be learned from the data nonparametrically. This model is equivalent to *only* assuming:

$$\log OR(w) = \beta_0 + \beta_1^T w.$$

Estimates and inference for the coefficient vector in this semiparametric model can be obtained by applying the R function *causalGLM* with the option 'estimand = "OR"'.

3.3 Conditional relative-risk (RR) and partially-linear log-linear regression

The partially-linear log-linear model is of the form:

$$\log\{E[Y|A, W]\} = \beta_0 + \beta_1^T W \cdot A + h_0(W),$$

where $h_0(W) := \log\{E[Y|A = 0, W]\}$ is an unspecified nuisance function that is to be learned from the data nonparametrically. This model is equivalent to *only* assuming:

$$\log RR(w) = \beta_0 + \beta_1^T w.$$

Estimates and inference for the coefficient vector in this semiparametric model can be obtained by applying the R function *causalGLM* with the option 'estimand = "RR"'.

3.4 causalGLM in practice

Let us generate a mock dataset that has a constant CATE of value 1.

```
> library(causalGLM)
> n <- 250
> W <- runif(n, min = -1, max = 1)
> A <- rbinom(n, size = 1, prob = plogis(W))
> Y <- rnorm(n, mean = A + W, sd = 0.3)
> data <- data.frame(W, A, Y)
```

We specify the parametric form of the CATE through the formula argument. In this case, we will use the intercept-only formula which is equivalent to assuming the CATE is constant. The output consists of: a coefficient estimate for the intercept, lower and upper confidence intervals, an asymptotic standard error estimate for the estimator, a Z-score and p-value. The argument 'W' should be a character vector of variable names in data for which to adjust, 'A' should be the name of a treatment variable, and 'Y' should be the name of an outcome variable. We set the argument 'estimand = "CATE"' to estimate the conditional average treatment effect.

```
> formula <- ~ 1
> output <-
+   causalGLMsp(
+     formula,
+     data,
+     W = "W", A = "A", Y = "Y",
+     estimand = "CATE",
+     verbose = FALSE
+   )
> summary(output)
```

	type	param	tmle_est	se	lower	upper	Z_score	p_value
1:	CATE (Intercept)		0.9660962	0.03617739	0.8951899	1.037003	422.2339	0

We can also do a much more complex model with higher dimensional treatment effect interactions. The following data distribution has $CATE(w) = 1 + w$.

```
> Y <-
+   rnorm(n,
+     mean = A * W + A + poly(W, degree = 3) + sin(4 * W),
+     sd = 0.4)
> data <- data.frame(W, A, Y)
> formula <- ~ 1 + W
> output <-
+   causalGLMsp(
+     formula,
+     data,
+     W = "W", A = "A", Y = "Y",
+     estimand = "CATE",
+     verbose = FALSE
+   )
> summary(output)
```

	type	param	tmle_est	se	lower	upper	Z_score	p_value
1:	CATE (Intercept)		1.0572233	0.05351641	0.9523330	1.162113	312.3559	0
2:	CATE	W	0.8594229	0.09597956	0.6713065	1.047539	141.5788	0

Currently, the nonparametric learning is performed using the partially-linear first-order Highly Adaptive Lasso (HAL) implemented using the R package `tlverse/hal9001`. HAL is an adaptive piece-wise linear regression spline estimator and the custom implementation performs the risk minimization entirely within the semiparametric model. It is implemented using LASSO regression with the parametric treatment interactions (as specified by the formula argument) unpenalized and a rich penalized spline basis for the nonparametric component of the model. This method performs risk minimization entirely within the semiparametric model.

There are a number of other built-in learning options: `c("HAL", "SuperLearner", "glm", "glmnet", "gam", "mars", "ranger", "xgboost")`

```
> Y <- rnorm(n, mean = A * W + A + W, sd = 0.4)
> data <- data.frame(W, A, Y)
```

```

> # generalized additive models:
> output <-
+   causalGLMsp(
+     formula,
+     data,
+     W = "W", A = "A", Y = "Y",
+     estimand = "CATE",
+     learning_method = "gam",
+     verbose = FALSE
+   )
> summary(output)

      type      param  tmle_est      se    lower    upper  Z_score p_value
1: CATE (Intercept) 0.9743861 0.05326979 0.8699792 1.078793 289.2145      0
2: CATE              W 1.0589704 0.10045665 0.8620790 1.255862 166.6768      0

> # multivariate adaptive regression splines:
> output <-
+   causalGLMsp(
+     formula,
+     data,
+     W = "W", A = "A", Y = "Y",
+     estimand = "CATE",
+     learning_method = "mars",
+     verbose = FALSE
+   )
> summary(output)

      type      param  tmle_est      se    lower    upper  Z_score p_value
1: CATE (Intercept) 0.9748477 0.05329876 0.8703841 1.079311 289.1943      0
2: CATE              W 1.0635533 0.09989848 0.8677558 1.259351 168.3334      0

> # gradient-boosting with xgboost: :
> output <-
+   causalGLMsp(
+     formula,
+     data,
+     W = "W", A = "A", Y = "Y",
+     estimand = "CATE",
+     learning_method = "xgboost",
+     verbose = FALSE
+   )
> summary(output)

      type      param  tmle_est      se    lower    upper  Z_score p_value
1: CATE (Intercept) 0.9719143 0.06324945 0.8479476 1.095881 242.9636      0
2: CATE              W 1.0523383 0.12225370 0.8127254 1.291951 136.1016      0

```

The default learning algorithm "HAL" can be customized with the `HAL_args_Y0W` argument (see the arguments in `hal9001` for more description). `'max_degree' = 1` corresponds with estimating the nuisance function $E[Y|A = 0, W]$ with an additive model. `'num_knots'` specifies for each interaction degree how many variable knot points are used to generate the tensor product interaction basis functions.

```

> HAL_args_Y0W <-
+   list(

```

```

+     smoothness_orders = 1,
+     max_degree = 2,
+     num_knots = c(10, 5, 1)
+   )
> output <-
+   causalGLMsp(
+     formula,
+     data,
+     W = "W", A = "A", Y = "Y",
+     estimand = "CATE",
+     learning_method = "HAL",
+     HAL_args_YOW = HAL_args_YOW,
+     verbose = FALSE
+   )
> summary(output)

```

	type	param	tmle_est	se	lower	upper	Z_score	p_value
1:	CATE (Intercept)		0.9751935	0.05277121	0.8717639	1.078623	292.1890	0
2:	CATE	W	1.0561517	0.09783143	0.8644056	1.247898	170.6939	0

It is also possible to employ custom learners using the tlverse/sl3 framework and the `sl3_Learner_Y` (to estimate $E[Y|A = 1, W]$ and $E[Y|A = 0, W]$) and `sl3_Learner_A` (to estimate $P(A = 1|W)$) argument. Take a look at the argument `append_interaction_matrix` to understand the design matrix that is given as input to the learner `sl3_Learner_Y`. In particular, it is important to note that `sl3_Learner_Y` will be sent to `Lrnr_glm_semiparametric`.

```

> library(sl3)
> lrnr_glmnet <- Lrnr_glmnet$new()
> lrnr_xgboost <- Lrnr_xgboost$new(max_depth = 4)
> lrnr_earth <- Lrnr_earth$new()
> lrnr_stack <-
+   make_learner(Stack, lrnr_glmnet, lrnr_xgboost, lrnr_earth)
> lrnr_cv <- Lrnr_cv$new(lrnr_stack, full_fit = TRUE)
> # A custom superlearner
> lrnr_sl <- make_learner(Pipeline, lrnr_cv, Lrnr_cv_selector$new())
> output <-
+   causalGLMsp(
+     formula,
+     data,
+     W = "W", A = "A", Y = "Y",
+     estimand = "CATE",
+     sl3_Learner_A = lrnr_sl ,
+     sl3_Learner_Y = lrnr_sl ,
+     verbose = FALSE
+   )
> summary(output)

```

	type	param	tmle_est	se	lower	upper	Z_score	p_value
1:	CATE (Intercept)		0.972336	0.05404197	0.8664157	1.078256	284.4823	0
2:	CATE	W	1.069745	0.10306803	0.8677350	1.271754	164.1066	0

That's all there is to it! `causalGLMsp` also supports the RR and OR estimands which are run in a completely analogous way. Use the option `'estimand = "OR"'` to estimate the conditional odds ratio, and use the option `'estimand = "rR"'` to estimate the conditional relative risk. Note that the parametric model specified by formula is actually for the log odds ratio and log relative risk (i.e. at the log scale). Thus, the coefficients returned are for the

log-transformed OR and RR. We also provide the exponential-transformed coefficients and confidence intervals, which may be more interpretable as measures of the OR and RR.

```
> n <- 250
> W <- runif(n, min = -1, max = 1)
> A <- rbinom(n, size = 1, prob = plogis(W))
> # OR
> Y <- rbinom(n, size = 1, prob = plogis(A + A * W + W + sin(5 * W)))
> data <- data.frame(W, A, Y)
> formula ~ 1 + W

formula ~ 1 + W

> output <-
+   causalGLMsp(
+     formula,
+     data,
+     W = "W", A = "A", Y = "Y",
+     estimand = "OR" ,
+     verbose = FALSE
+   )
> summary(output)

      type      param  tmle_est      se      lower      upper  psi_exp lower_exp
1:   OR (Intercept) 0.9595055 0.3067406 0.3583050 1.560706 2.610405 1.4309020
2:   OR              W 0.4888635 0.5106881 -0.5120668 1.489794 1.630462 0.5992558
      upper_exp  Z_score p_value
1:  4.762182 49.45911      0
2:  4.436180 15.13568      0

> # RR
> Y <- rpois(n, lambda = exp(A + A * W + sin(5 * W)))
> data <- data.frame(W, A, Y)
> formula ~ 1 + W

formula ~ 1 + W

> output <-
+   causalGLMsp(
+     formula,
+     data,
+     W = "W", A = "A", Y = "Y",
+     estimand = "RR" ,
+     verbose = FALSE
+   )
> summary(output)

      type      param  tmle_est      se      lower      upper  psi_exp lower_exp
1:   RR (Intercept) 0.9620825 0.09574633 0.7744231 1.149742 2.617141 2.169340
2:   RR              W 1.0155014 0.19284446 0.6375332 1.393470 2.760747 1.891808
      upper_exp  Z_score p_value
1:  3.157378 158.87668      0
2:  4.028804 83.26133      0

>
```

4 Robust nonparametric generalized linear models for treatment effect estimation with causalGLMnp

In the previous section, we considered semiparametric generalized linear models where the data distribution component of interest (the estimand) is modeled parametrically and nuisance components are modeled nonparametrically. While this is much more robust than typical parametric methods, the parametric assumption on the estimand of interest can still be quite strong. It is therefore of interest to develop fully nonparametric methods that provide correct and interpretable estimates and inference under no parametric assumptions. To do so, we will still utilize user-specified parametric models for the estimand of interest, however, we will not assume that these parametric models are correct. We will treat these parametric models as interpretable approximations of the true nonparametric estimand. That is, we utilize the parametric model as a "working-model" that is only used to derive an interesting nonparametric estimand.

It turns out that many of these working models have desirable properties. In particular, by specifying parametric models/formulas that depend on $V \subset W$ where V is a subvector of baseline covariates, we can actually learn marginal structural models. Notably, the intercept working model that approximates the true estimand by a constant often corresponds with a marginal causal estimand like the average treatment effect (ATE or ATT), the marginal treatment-specific mean, or the marginal relative risk. For these reasons, these nonparametrics method allow for the estimation of an even more rich class of parameters than the analogous semiparametric methods.

Nonparametric working-model based estimators for the estimands are implemented in the function 'causalGLMnp'. These methods have:

- Interpretable coefficient-based estimates and inference
- Nonparametric and causal estimates and inference even when the parametric model is incorrect
- Many of the estimands correspond with marginal structural models when lower dimensional working-models are used.

4.1 Conditional average treatment effect (CATE) estimation with a linear working-model

To define a nonparametric approximation of the true CATE with a parametric linear-working model, we utilize the least-squares projection. Let $\underline{f}(w)$ be an arbitrary known vector-valued function of the covariates and consider the linear parametric working-model $\beta^T \underline{f}(w)$ for $CATE(w)$.

Define the risk function,

$$R_{CATE}(\beta) = E \left\{ CATE(W) - \beta^T \underline{f}(W) \right\}^2$$

Our estimand of interest is given by (β_0^*, β_1^*) which is defined as the minimizer of R_{CATE} .

Consider the simple working model $\beta^T \underline{f}(W) := \beta_0 + \beta_1^T W$. The risk function then reduces to

$$R_{CATE}(\beta_0, \beta_1) = E \left\{ CATE(W) - \beta_0 - \beta_1^T W \right\}^2$$

which can be viewed as the ordinary least-squares regression of the true estimand $CATE(W)$ onto W .

By specifying a lower dimensional working model $\underline{f}(W) := V$ for some $V \subset W$, the risk function further reduces to

$$R_{CATE}(\beta_0, \beta_1) = E \left\{ E[CATE(W)|V] - \beta_0 - \beta_1^T V \right\}^2.$$

The risk minimizer is now the least-squares projection of the true marginal structural CATE model $E[CATE(W)|V]$ onto the linear working model. Note if $E[CATE(W)|V] = \beta_0 + \beta_1^T V$, so that the working model is correct, then this estimand can be directly interpreted as a marginal structural CATE function.

If we use the intercept model then the risk function reduces to

$$R_{CATE}(\beta_0) = E \left\{ E[CATE(W)] - \beta_0 \right\}^2.$$

and the risk minimizer is exactly given by the nonparametric ATE $E[CATE(W)]$! Thus, the intercept model can be used for marginal ATE estimation.

Quite remarkably, this estimands based on such least-squares working model projections automatically reduce to marginal structural model parameters when lower dimensional working models are used. No user or developer intervention is needed for this to happen!

4.2 Conditional average treatment effect among the treated (CATT) estimation with a linear working-model

We now define an alternative working model for the CATE that focuses on the treatment effect among the treated. Again, let $\underline{f}(w)$ be an arbitrary known vector-valued function of the covariates and consider the linear parametric working-model $\beta^T \underline{f}(w)$ for $CATE(w)$.

Define the risk function,

$$R_{CATT}(\beta) = E \left\{ E[Y|A, W] - A \cdot \beta^T \underline{f}(W) - E[Y|A = 0, W] \right\}^2$$

Our estimand of interest is β^* which is defined as the minimizer of R_{CATT} . This working model can be viewed as the least-squares regression of the true conditional mean $E[Y|A, W]$ onto the interaction model $A \cdot \beta^T \underline{f}(W)$ using as offset the true placebo conditional mean $E[Y|A = 0, W]$. It turns out that we can rewrite this risk function as

$$R_{CATT}(\beta) = E \left\{ A \left[CATE(W) - \beta^T \underline{f}(W) \right] \right\}^2,$$

which is the least-squares projection of the true CATE onto the linear working model using only the observations with $A = 1$ (the treated). Because of this, we call estimands based on this risk function measures of the conditional average treatment effect among the treated (CATT).

This method can also be used to estimate marginal structural models for treatment effects among the treated.

By specifying a lower dimensional working model $\underline{f}(W) := V$ for some $V \subset W$, the risk function further reduces to

$$R_{CATT}(\beta_0, \beta_1) = E \left\{ A \left[E[CATE(W)|V, A = 1] - \beta_0 - \beta_1^T V \right] \right\}^2.$$

The risk minimizer is now the least-squares projection of the true marginal structural CATT model $E[CATE(W)|V, A = 1]$ onto the linear working model.

If we use the intercept model then the risk function reduces to

$$R_{CATT}(\beta_0) = E \left\{ A \cdot [E[CATE(W)|A = 1] - \beta_0]^2 \right\}.$$

and the risk minimizer is exactly given by the nonparametric ATT $E[CATE(W)|A = 1]$! Thus, the intercept model can be used for marginal ATT estimation.

4.3 Conditional treatment-specific mean (TSM) estimation with a linear working-model

A similar working model-based estimand can be constructed for the conditional treatment specific mean. Let a be a level of a categorical treatment assignment A . Again, let $\underline{f}(w)$ be an arbitrary known vector-valued function of the covariates and consider the linear parametric working-model $\beta^T \underline{f}(w)$ for $CATE(w)$.

Define the risk function,

$$R_{TSM}(\beta) = E \left\{ E[Y|A = a, W] - \beta^T \underline{f}(W) \right\}^2$$

Our estimand of interest is β^* which is defined as the minimizer of R_{TSM} . This estimand corresponds with the least-squares projection of $E[Y|A = a, W]$ onto the linear working model.

By specifying a lower dimensional working model $\underline{f}(W) := V$ for some $V \subset W$, the risk function further reduces to

$$R_{TSM}(\beta_0, \beta_1) = E \left\{ E[E[Y|A=a, W]|V] - \beta_0 - \beta_1^T V \right\}^2.$$

The risk minimizer is now the least-squares projection of the true marginal structural TSM model $E[E[Y|A=a, W]|V]$ onto the linear working model.

If we use the intercept model then the risk function reduces to

$$R_{TSM}(\beta_0) = E \{ E[E[Y|A=a, W]] - \beta_0 \}^2.$$

and the risk minimizer is exactly given by the nonparametric TSM $E[E[Y|A=a, W]]$! Thus, the intercept model can be used for marginal TSM estimation.

4.4 Conditional odds-ratio (OR) estimation with a logistic working-model

Define the working model

$$P_\beta(Y = 1|A, W) := \text{expit} \left\{ A \cdot \beta^T \underline{f}(W) + \text{logit}(P(Y = 0|A, W)) \right\},$$

which is not assumed correct.

Consider the log-likelihood projection risk function

$$R_{OR}(\beta) = E \left\{ P(Y = 1|A, W) \log(P_\beta(Y = 1|A, W)) + P(Y = 0|A, W) \log(P_\beta(Y = 0|A, W)) \right\}.$$

We define the nonparametric OR estimand as the risk minimizer β^* of R_{OR} . This estimand unfortunately does not reduce to a marginal structural model estimand when $\underline{f}(W)$ lower dimensional.

4.5 Conditional relative-risk regression (RR) with a log-linear working-model

Define the log-linear multiplicative working model

$$E_\beta[Y|A = 1, W] := \exp \left\{ \beta^T \underline{f}(W) \right\} E[Y|A = 0, W],$$

which is not assumed correct and $E_\beta[Y|A = 0, W] := E[Y|A = 0, W]$ is left correctly specified. We define the projection using the log-linear generalized linear model,

$$R_{RR}(\beta) = E \left\{ E[Y|A = 0, W] \exp \left\{ \beta^T \underline{f}(W) \right\} - E[Y|A = 1, W] \beta^T \underline{f}(W) \right\}.$$

We define the nonparametric RR estimand as the risk minimizer β^* of R_{RR} .

By specifying a lower dimensional working model $\underline{f}(W) := V$ for some $V \subset W$, the risk function further reduces to

$$R_{RR}(\beta) = E \left\{ E[E[Y|A = 0, W]|V] \exp \left\{ \beta^T V \right\} - E[E[Y|A = 1, W]|V] \beta^T V \right\}.$$

The risk minimizer is now the projection of the true marginal structural RR model $\frac{E[E[Y|A=1, W]|V]}{E[E[Y|A=0, W]|V]}$ onto the log-linear working model. Thus, if the working model is correctly specified, the estimand is $\frac{E[E[Y|A=1, W]|V]}{E[E[Y|A=0, W]|V]}$.

If we use the intercept model then the risk function reduces to

$$R_{RR}(\beta) = E \{ E[E[Y|A = 0, W]] \exp \{ \beta \} - E[E[Y|A = 1, W]] \beta \}.$$

and the risk minimizer is exactly given by the nonparametric marginal relative risk $\frac{E[E[Y|A=1, W]]}{E[E[Y|A=0, W]]}$! Thus, the intercept model can be used for marginal RR estimation.

5 High dimensional semiparametric generalized linear models for treatment effect estimation using the LASSO with causalGLMnet

6 Robust nonparametric inference for the hazard ratio with npCOXph