A guided tour in targeted learning territory

David Benkeser, Antoine Chambaz, Nima Hejazi

10/11/2017

Contents

This is a very first draft of our article. The current *tentative* title is

A guided tour in targeted learning territory

```
set.seed(54321) ## because reproducibility matters...
suppressMessages(library(R.utils)) ## make sure it is installed
expit <- plogis
logit <- qlogis
```

1 Introduction

We are interested in a reproducible experiment. The generic summary of how one realization of the experiment unfolds, our observation, is called O. We view O as a random variable drawn from what we call the law P_0 of the experiment. The law P_0 is viewed as an element of what we call the model. Denoted by $\mathcal{M}$, the model is the collection of *all* laws from which O can be drawn. The more we know about the experiment, the smaller is $\mathcal{M}$. In all our examples, model $\mathcal{M}$ will put no restriction whatsoever on the candidate laws.

Consider the following chunk of code:

```
drawFromExperiment <- function(n, intervention = c("none", "zero", "one")) {
  ##
  ## preliminary
  ##
  n <- Arguments$getInteger(n, c(1, Inf))
  intervention <- match.arg(intervention)
  ## 'gbar' and 'Qbar' factors
  gbar <- function(W) {
    expit(-0.3 + 2 * W - 1.5 * W^2)
  }
  Qbar <- function(AW) {
    A <- AW[, 1]
    W <- AW[, 2]
    A * cos(2 * pi * W) + (1 - A) * sin(2 * pi * W^2)
  }
  ##
  ## sampling
  ##

  ## context
  W <- runif(n)

  ## counterfactual rewards
  zeroW <- cbind(A = 0, W)
```

```
  oneW <- cbind(A = 1, W)
  Yzero <- rnorm(n, mean = Qbar(zeroW), sd = 1)
  Yone <- rnorm(n, mean = Qbar(oneW), sd = 1)

  ## action undertaken
  if (intervention == "none") {
    A <- rbinom(n, size = 1, prob = gbar(W))
  } else if (intervention == "zero") {
    A <- rep(0, n)
  } else if (intervention == "one") {
    A <- rep(1, n)
  }

  ## actual rewards
  Y <- A * Yone + (1 - A) * Yzero

  ## observation
  obs <- cbind(W = W, A = A, Y = Y)
  attr(obs, "gbar") <- gbar
  attr(obs, "Qbar") <- Qbar

  return(obs)
}
```

We can interpret `drawFromExperiment` as a law P_0 since we can use the function to sample observations from a common law. It is even a little more than that, because we can **intervene** on the experiment, by setting its `intervention` argument to either `"zero"` or `"one"` (the default value, `"none"`, corresponds to the absence of intervention). The next chunk of code runs the experiment five times independently:

```
obs.five <- drawFromExperiment(5)
obs.five
```

```
##              W A          Y
## [1,] 0.4290078 1 -0.4410530
## [2,] 0.4984304 0  0.1905250
## [3,] 0.1766923 0 -0.1325567
## [4,] 0.2743935 1 -1.6415356
## [5,] 0.2165102 0  0.6829717
## attr(,"gbar")
## function (W)
## {
##     expit(-0.3 + 2 * W - 1.5 * W^2)
## }
## <bytecode: 0x408c458>
## <environment: 0x4ca7838>
## attr(,"Qbar")
## function (AW)
## {
##     A <- AW[, 1]
##     W <- AW[, 2]
##     A * cos(2 * pi * W) + (1 - A) * sin(2 * pi * W^2)
## }
## <bytecode: 0x3c1b608>
## <environment: 0x4ca7838>
```

The `attributes` of the object `obs` are visible because we act as oracles, *i.e.*, we know completely the nature of the experiment. From a probabilistic point of view, the attributes `gbar` and `Qbar` are infinite-dimensional features of P_0. There is more to P_0 than $\bar{g}_0$ and $\bar{Q}_0$, formally defined by

$$\bar{g}_0(W) \equiv P_0(A = 1|W), \quad \bar{Q}_0(A, W) \equiv E_{P_0}(Y|A, W), \tag{1}$$

for instance the marginal distribution of W under P_0, and the conditional distribution (not expectation) of Y given (A, W), but $\bar{g}_0$ and $\bar{Q}_0$ will play a prominent role in our story.