

O/E ratio

Jasper van Egeraat

2026-02-17

Counterfactual O/E ratio

Let M be a model that estimates risk under some treatment a and given some covariates.

The counterfactual observed/expected (O/E) ratio under treatment a is the outcome proportion divided by the mean predicted risk in a population where we would (counterfactually) have treated everyone with treatment a .

In observational data, parts of the population did not receive treatment a . For these people we do not observe their (counterfactual) outcome under treatment a . Thus, we have to do some IPTW tricks to form a pseudopopulation that represents the original data if we had treated every individual with treatment a .

We can however estimate patients' risk under some treatment a using our model M , regardless of patients treatment status. We do not need to form a pseudopopulation.

To me, it seemed now reasonable that the O/E ratio can be estimated in two ways, which I thought to be equivalent. For both ways, the numerator ('observed') is estimated as follows

$E_{ps}[\text{observed}] = \text{weighted.mean}(\text{pseudo}\$outcome, \text{pseudo}\$iptw),$

where E_{ps} denotes the expected value in the IPTW pseudopopulation, and $\text{pseudo}\$$ is the subset of patients that *were* treated with treatment a .

The denominator ('expected') can be estimated in two ways:

1. $E[\text{risk}] = \text{mean}(\text{data}\$risk)$
2. $E_{ps}[\text{risk}] = \text{weighted.mean}(\text{pseudo}\$risk, \text{pseudo}\$iptw)$

where $E[\text{risk}]$ is the estimated risk (from model M) in the original full population $\text{data}\$$.

In many cases these two methods seemed equivalent and computed the same result. However, in this example it can be shown they are not, and we see which definition should be used. Henceforth, I will call the O/E ratio based on option 1 *oeratio*, and the ratio based on option 2 *oeratio_pp*.

```
library(CFeval)

n <- 10000

data <- data.frame(L = rnorm(n))
data$A <- rbinom(n, 1, plogis(data$L))
data$Y0 <- rbinom(n, 1, plogis(0.1 + 0.5*data$L))
data$Y1 <- rbinom(n, 1, plogis(0.1 + 0.5*data$L - 4*data$A))
data$Y <- ifelse(data$A == 1, data$Y1, data$Y0)
```

A simple dataframe, where the effect of treatment A on outcome Y is confounded by L . $Y0$ and $Y1$ are the (unobserved) outcomes under treatments 0 and 1. Y is the observed outcome.

We are going to develop and validate some models on this dataset. One model gives random predictions, one model 100% accurately predicts the observed outcome Y , and one 100% accurately predicts the counterfactual outcome under treatment 0.

```
random_predictions <- runif(n, 0, 1)
naive_perfect <- data$Y
causal_perfect <- data$Y0
```

How well calibrated should each model be in a world where we would have treated nobody? The model that perfectly predicts Y_0 should have O/E ratio equal to 1.0. The naive model should not be well-calibrated, because it predicts the outcome under some prespecified treatment assignment strategy, whereas we are estimating the performance of the model in a population that follows a different treatment assignment strategy (nobody gets treatment).

```
CFscore(
  object = list(
    "random" = random_predictions,
    "naive" = naive_perfect,
    "causal" = causal_perfect
  ),
  data = data,
  outcome_formula = Y ~ 1,
  treatment_formula = A ~ L,
  treatment_of_interest = 0,
  metrics = c("oeratio", "oeratio_pp")
)
```

```
## Estimation of the performance of the prediction model in a
## counterfactual (CF) dataset where everyone's treatment A was set to 0.
## The following assumptions must be satisfied for correct inference:
## - Conditional exchangeability requires that given IP-weights are
## sufficient to adjust for confounding and selection bias between
## treatment and outcome.
## - Positivity (assess $weights for outliers)
## - Consistency
## - No interference
## - Correctly specified propensity formula. Estimated treatment model is
## logit(A) = 0.01 + 0.97*L
##
##      model oeratio oeratio_pp
## null model  1.000      1.00
##   random  1.038      1.04
##    naive  2.082      1.00
##   causal  0.998      1.00
```

Note the big difference between *oeratio* and *oeratio_pp* for the naive predictions.

TODO: Volgens past hier een goede uitleg over wanneer en waarom *oeratio* en *oeratio_pp* soms wel verschillen en soms niet.

Thus, we conclude that to compute the counterfactual O/E ratio, we should not use the estimated risks in the weighted pseudopopulation but rather the mean estimated risk in the original population.