Possible issue of the knockoff R package

The fixed-X feature statistics `stat.glmnet_lambdasmax` (and others) in the R package implementation employs a `glmnet` Lasso solver with `intercept=TRUE` by default (line 177 and 206).

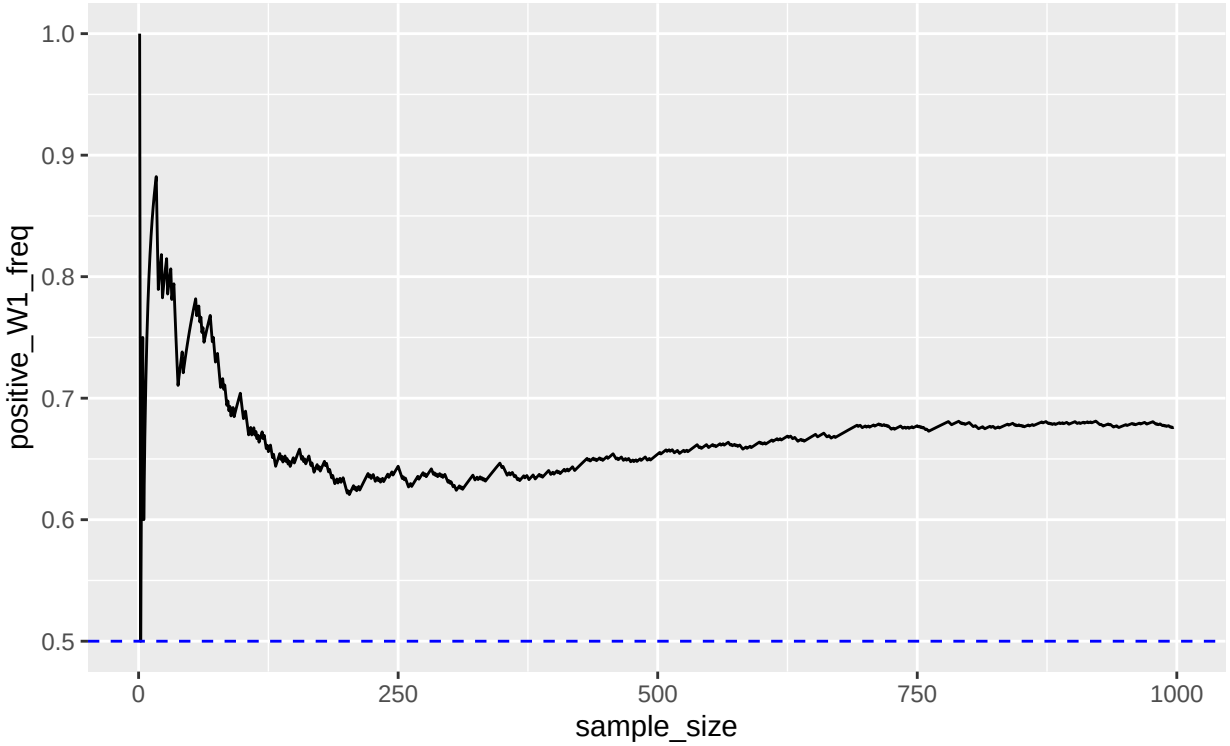
However, the knockoff matrix is not adjusted for this intercept term during construction, i.e. we don't have $X^T\mathbf{1} = \tilde{X}^T\mathbf{1}$ with $\mathbf{1}$ being the vector of ones. Would this fails the sufficiency principle and further the validity of the method?

Examples

Here is an example showing the null variable X_1 has $\text{sgn}(W_1)$ not Rademacher.

Null W-stat doesn't have Rademacher sign

```
set.seed(2022)

# X is a 2-by-n matrix
n <- 10

X1 <- c(0, rep(1, n-1))  # X1 highly correlated with rep(1, n)
X2 <- c(rep(0, n/2), rep(1, n/2))  # X2 correlated with X1 and rep(1, n)
X <- cbind(X1, X2)

beta <- c(0, 0)  # all variables are null

kn_vars <- create.fixed(X) # create knockoff matrix

# run experiment
W_signs <- replicate(1000, {
  # generate y
  y <- X %*% beta + rnorm(n)
  # run the feature statistics function
  W <- stat.glmnet_lambdasmax(kn_vars$X, kn_vars$Xk, y)
  return(sign(W[1]))  # sign of W_1, expected to be Unif{+1, -1}
})


W_signs <- W_signs[W_signs != 0] # neglect the cases where W_1 = 0
sample_num <- length(W_signs)


# plot figure
plot_data <- data.frame(sample_size = 1:sample_num,
                        positive_W1_freq = cumsum(W_signs>0)/(1:sample_num))
ggplot(plot_data) +
  geom_line(aes(x = sample_size, y = positive_W1_freq)) +
  geom_hline(yintercept = 0.5, color = "blue", linetype = "dashed")
```



FDR is not controlled

FDR may also not be controlled at the desired level (0.5 in this example) due to the same reason.

The problem setting is a bit different from the previous one and can be found in the Rmd file.