

Estimating Variance of Simple Defined Variable Effect directly

Felix Kapulla

```
knitr::opts_chunk$set(fig.width=13, fig.height=7)

rm(list=ls())
library(Matrix)
library(tidyverse)
library(ggplot2)
library(ggpubr)
library(ranger)
library(MixMatrix)
library(mvtnorm)
library(stringr)
library(parallel)

source('C:/Users/feix_/iCloudDrive/Studium Master/CQM - Thesis Internship/Thesis-VariableEffects/Baselin

cores <- detectCores()
clust <- makeCluster(cores)
parallel::clusterEvalQ(clust,
                        expr = {source('C:/Users/feix_/iCloudDrive/Studium Master/CQM - Thesis Internship
```

Simulation

```
##### Simulation Setup
n <- c(1000, 10000) ; num.trees <- 500 ; repeats <- 200; cor <- c(0, 0.8); k <- c(1, 0.3)
formulas <- c("2*x.1+4*x.2-0.5*x.3") #" -0.5*x.1^3+3*x.2+0.5*sqrt(abs(x.3))"
scenarios <- data.frame(expand.grid(n, num.trees, formulas, repeats, cor, k))
colnames(scenarios) = c("N", "N_Trees", "Formula", "Repeats", "Correlation", "k")
scenarios[, "Formula"] <- as.character(scenarios[, "Formula"]) ### Formula became Factor
scenarios <- split(scenarios, seq(nrow(scenarios)))

#system.time(result <- lapply(X = scenarios, FUN = sim_multi))
#Run Simulation
system.time(result <- parLapply(cl = clust,
                              X = scenarios,
                              fun = sim_multi))

## user system elapsed
## 0.02 0.19 3871.89
```

```
print_results(result)
```

```
## Setting: N = 1000 ; k = 1 ; Correlation = 0 ;
##      Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 500
## Mean(s) of simulated RF Variable Effect(s):
##      1.890015 3.909591 -0.4172731
## Mean(s) of simulated LM Variable Effect(s):
##      1.999495 4.002157 -0.4955707
## Mean(s) of simulated True Formula Variable Effect(s):
##      2 4 -0.5
## Mean(s) of True Variable Effect(s):
##      2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
##      0.2388217 0.2504319 0.2128206 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      0.3201844 0.3244954 0.3125216
##
## Setting: N = 10000 ; k = 1 ; Correlation = 0 ;
##      Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 500
## Mean(s) of simulated RF Variable Effect(s):
##      1.964404 3.996436 -0.4783074
## Mean(s) of simulated LM Variable Effect(s):
##      2.000878 3.999523 -0.5013587
## Mean(s) of simulated True Formula Variable Effect(s):
##      2 4 -0.5
## Mean(s) of True Variable Effect(s):
##      2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
##      0.2000501 0.1994035 0.1858077 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      0.4897796 0.5136054 0.5331296
##
## Setting: N = 1000 ; k = 1 ; Correlation = 0.8 ;
##      Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 500
## Mean(s) of simulated RF Variable Effect(s):
##      1.874866 3.527891 -0.04552301
## Mean(s) of simulated LM Variable Effect(s):
##      2.001779 4.000792 -0.5004559
## Mean(s) of simulated True Formula Variable Effect(s):
##      2 4 -0.5
## Mean(s) of True Variable Effect(s):
##      2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
##      0.2393762 0.2338161 0.2535034 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##      0.3148377 0.3041681 0.2899013
##
## Setting: N = 10000 ; k = 1 ; Correlation = 0.8 ;
##      Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 500
## Mean(s) of simulated RF Variable Effect(s):
##      1.946222 3.819122 -0.374218
## Mean(s) of simulated LM Variable Effect(s):
##      1.998696 4.000206 -0.501528
```

```

## Mean(s) of simulated True Formula Variable Effect(s):
## 2 4 -0.5
## Mean(s) of True Variable Effect(s):
## 2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
## 0.1886302 0.2100276 0.193163 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.4635709 0.4708958 0.4624211
##
## Setting: N = 1000 ; k = 0.3 ; Correlation = 0 ;
## Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 500
## Mean(s) of simulated RF Variable Effect(s):
## 2.088646 4.380539 -0.4224318
## Mean(s) of simulated LM Variable Effect(s):
## 2.001559 4.003015 -0.5014391
## Mean(s) of simulated True Formula Variable Effect(s):
## 2 4 -0.5
## Mean(s) of True Variable Effect(s):
## 2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
## 0.6828274 0.836296 0.5013946 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.9980357 1.027398 0.9324713
##
## Setting: N = 10000 ; k = 0.3 ; Correlation = 0 ;
## Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 500
## Mean(s) of simulated RF Variable Effect(s):
## 2.146615 4.135051 -0.5088369
## Mean(s) of simulated LM Variable Effect(s):
## 1.998169 4.000439 -0.4996974
## Mean(s) of simulated True Formula Variable Effect(s):
## 2 4 -0.5
## Mean(s) of True Variable Effect(s):
## 2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
## 0.6367413 0.637573 0.5089145 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 1.423285 1.581029 1.521695
##
## Setting: N = 1000 ; k = 0.3 ; Correlation = 0.8 ;
## Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 500
## Mean(s) of simulated RF Variable Effect(s):
## 1.918711 3.910479 -0.2892278
## Mean(s) of simulated LM Variable Effect(s):
## 2.000738 4.001504 -0.5023809
## Mean(s) of simulated True Formula Variable Effect(s):
## 2 4 -0.5
## Mean(s) of True Variable Effect(s):
## 2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
## 0.6449845 0.6331564 0.5319982 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 0.8134407 0.8779091 0.7202576
##

```

```
## Setting: N = 10000 ; k = 0.3 ; Correlation = 0.8 ;
##      Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 500
## Mean(s) of simulated RF Variable Effect(s):
## 1.977346 4.001932 -0.4955097
## Mean(s) of simulated LM Variable Effect(s):
## 1.997918 4.00254 -0.5000928
## Mean(s) of simulated True Formula Variable Effect(s):
## 2 4 -0.5
## Mean(s) of True Variable Effect(s):
## 2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
## 0.6293457 0.6127667 0.6006492 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
## 1.353877 1.21849 1.255176
```

```
result_plots <- plot_results(result)
result_plots
```

