Estimating Variance of Simple Defined Variable Effect directly

Felix Kapulla

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

```
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/Baseli

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

Simulation

```
###### Simulation Setup
n <- c(1000, 10000) ; num.trees <- 2000 ; repeats <- 100; 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
##     0.19     0.69 18179.70
```

```
print_results(result)
```

```
## Setting: N = 1000 ; k = 1 ; Correlation = 0 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.898973 3.832386 -0.4615647
## Mean(s) of simulated LM Variable Effect(s):
##   1.999355 3.997691 -0.4997978
## Mean(s) of True Variable Effect(s):
##   2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
##   0.2307777 0.2258448 0.2336837 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2488864 0.2390155 0.2234679
##
## Setting: N = 10000 ; k = 1 ; Correlation = 0 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.963991 3.993768 -0.4954206
## Mean(s) of simulated LM Variable Effect(s):
##   2.000833 3.99932 -0.5006134
## Mean(s) of True Variable Effect(s):
##   2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
##   0.2032975 0.2128785 0.1979169 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2678704 0.2397374 0.2483661
##
## Setting: N = 1000 ; k = 1 ; Correlation = 0.8 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.851594 3.515599 -0.08777112
## Mean(s) of simulated LM Variable Effect(s):
##   2.00532 3.994538 -0.5032325
## Mean(s) of True Variable Effect(s):
##   2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
##   0.2207407 0.2169502 0.2495088 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2216054 0.2391274 0.2200809
##
## Setting: N = 10000 ; k = 1 ; Correlation = 0.8 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.940592 3.822576 -0.3475832
## Mean(s) of simulated LM Variable Effect(s):
##   1.99625 3.999943 -0.4979572
## Mean(s) of True Variable Effect(s):
##   2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
##   0.1879126 0.1933718 0.1777165 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.2447607 0.2502332 0.2456043
```

```
##
## Setting: N = 1000 ; k = 0.3 ; Correlation = 0 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   2.010713 4.300178 -0.3984028
## Mean(s) of simulated LM Variable Effect(s):
##   1.99911 3.99918 -0.5002835
## Mean(s) of True Variable Effect(s):
##   2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
##   0.7656584 0.8579738 0.5250982 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.6591632 0.7985492 0.5765821
##
## Setting: N = 10000 ; k = 0.3 ; Correlation = 0 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.977679 4.17357 -0.603747
## Mean(s) of simulated LM Variable Effect(s):
##   2.001433 4.001318 -0.5002216
## Mean(s) of True Variable Effect(s):
##   2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
##   0.6213463 0.6275799 0.5038182 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.6777997 0.7121212 0.6504175
##
## Setting: N = 1000 ; k = 0.3 ; Correlation = 0.8 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   1.97552 3.855898 -0.3054135
## Mean(s) of simulated LM Variable Effect(s):
##   2.001595 3.999067 -0.5016144
## Mean(s) of True Variable Effect(s):
##   2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
##   0.6687945 0.7258811 0.5891783 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.6380022 0.6986885 0.5670916
##
## Setting: N = 10000 ; k = 0.3 ; Correlation = 0.8 ;
##          Formula = 2*x.1+4*x.2-0.5*x.3 ; N_Trees = 2000
## Mean(s) of simulated RF Variable Effect(s):
##   2.015982 3.983229 -0.4264835
## Mean(s) of simulated LM Variable Effect(s):
##   2.000005 4.000581 -0.499636
## Mean(s) of True Variable Effect(s):
##   2 4 -0.5
## Standard Error of simulated Variable Effects (RF):
##   0.6033368 0.6242727 0.6219175 .
## Mean of Standard Errors Estimates of Variable Effects (RF):
##   0.6934165 0.714088 0.677619
```

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