Bootstrap of Subjects Demonstration

Edward Gunning

2023-07-13

A very quick check/demonstration of the bootstrap of subjects function.

Load packages and functions:

```
set.seed(123)
library(splines)
functions_path <- here::here("code", "functions")
source(file.path(functions_path, "generate-basis-coefficient-matrix.R"))
source(file.path(functions_path, "add_natural_splines_to_df.R"))
source(file.path(functions_path, "add_poly_to_df.R"))
source(file.path(functions_path, "fit_spline_subject_ri_side.R"))
source(file.path(functions_path, "fit_naive_spline_intercept.R"))
source(file.path(functions_path, "generate_design.R"))
source(file.path(functions_path, "generate_polynomial_model_basis_coefficient.R"))
source(file.path(functions_path, "bootstrap_of_subjects.R"))
source(file.path(functions_path, "extract_fixef_coef.R"))
```

Generate data:

```
N_test <- 284
K <- 1
n_i_test <- 80
design_df <- generate_design_multiple_subjects(N = N_test,
                                               n_i = n_i_test,
                                               speed_sd = 1.5)
design_df <- add_poly_to_df(df = design_df,
                            poly_object = poly(x = seq(0, 1, length.out = 101),
                                                                degree = 2,
                                                                raw = FALSE))
sim_param_final <- readRDS(
  file = here::here("outputs",
                    "simulation",
                    "simulation-parameters-final.rds"))

Y_star_df <- generate_basis_coefficient_matrix(
  design_df = design_df,
  N = N_test,
  K = K,
  beta_poly_1_vec = sim_param_final$beta_poly_1_true[seq_len(K)],
  beta_poly_2_vec = sim_param_final$beta_poly_2_true[seq_len(K)],
  beta_poly_3_vec = sim_param_final$beta_poly_3_true[seq_len(K)],
  beta_sex_vec = sim_param_final$beta_sex_true[seq_len(K)],
  beta_speed_cent_vec = sim_param_final$beta_speed_cent_true[seq_len(K)],
  Q_star_array = sim_param_final$Q_star_true[,,seq_len(K), drop = FALSE],
```

```
  R_star_array =  sim_param_final$R_star_true[,,seq_len(K), drop = FALSE],
  s_vec = sim_param_final$s_k_true[seq_len(K)])
```

```
## Loading required package: mvtnorm
```

Do initial fit:

```
test_fit <- fit_spline_subject_ri_side(df_scores = Y_star_df,
                                       K_retain = K,
                                       df = 3,
                                       diagonal_covariance = FALSE)
```

Do bootstrap:

```
n_cores <- parallel::detectCores() - 1
B <- 250
boot_time <- system.time(boot_results <- bootstrap_of_subjects(
  df_for_bootstrap = Y_star_df,
  k_retain = K,
  model = "spline_subject_ri_side",
  df = 3,
  diagonal_covariance = FALSE,
  B = B,
  par_mc = TRUE,
  n_cores = n_cores))

print(paste0(100 *  mean(sapply(boot_results, function(x) {x[["singular"]]})),
             "% of fits were singular!"))
```

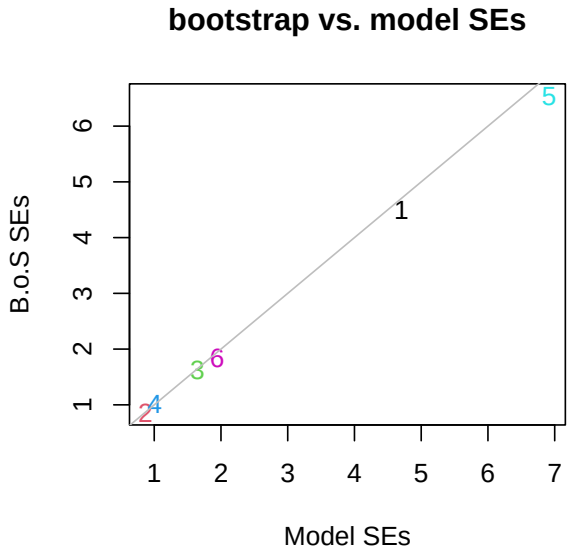
```
## [1] "28.8% of fits were singular!"
```

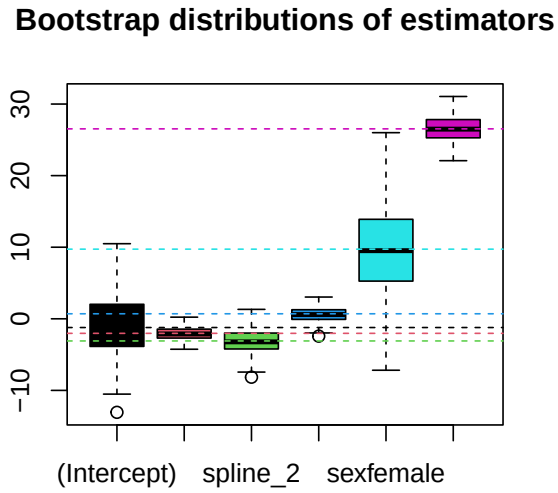
```
print(paste("bootstrap took", round(boot_time["elapsed"]/60,2), "mins"))
```

```
## [1] "bootstrap took 31.28 mins"
```

```
score_1_results <- sapply(boot_results, function(x) {x$fixef[,1]})
par(mfrow = c(1, 2))
plot(x = sqrt(diag(vcov(test_fit$lme_fit_list[[1]]))),
     y = apply(score_1_results, 1, sd),
     col = 1:6,
     pch = paste0(1:6),
     xlab = "Model SEs",
     ylab = "B.o.S SEs")
abline(0, 1, col = "grey")
title("bootstrap vs. model SEs")

boxplot(t(score_1_results), col = 1:6)
abline(h = fixef(test_fit$lme_fit_list[[1]]), col = 1:6, lty = 2)
title("Bootstrap distributions of estimators")
```






```
rnorm(n = 1)
```

```
## [1] -0.04686628
```

```
sessionInfo()
```

```
## R version 4.1.2 (2021-11-01)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_IE.UTF-8/en_IE.UTF-8/en_IE.UTF-8/C/en_IE.UTF-8/en_IE.UTF-8
##
## attached base packages:
## [1] parallel  splines   stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
## [1] mvtnorm_1.1-3  lme4_1.1-30    Matrix_1.4-0   progress_1.2.2
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_1.0.10       highr_0.9         pillar_1.7.0      compiler_4.1.2
##  [5] nloptr_2.0.0      prettyunits_1.1.1 tools_4.1.2       boot_1.3-28
##  [9] digest_0.6.29     tibble_3.1.6      evaluate_0.15     lifecycle_1.0.3
## [13] nlme_3.1-155      lattice_0.20-45   pkgconfig_2.0.3   rlang_1.1.1
## [17] DBI_1.1.2         cli_3.6.0         rstudioapi_0.13   yaml_2.3.5
## [21] xfun_0.29         fastmap_1.1.0     stringr_1.4.0     dplyr_1.0.8
## [25] knitr_1.37        generics_0.1.2    vctrs_0.5.1       hms_1.1.1
## [29] tidyselect_1.1.1  rprojroot_2.0.2   grid_4.1.2        glue_1.6.2
## [33] here_1.0.1        R6_2.5.1          fansi_1.0.2       rmarkdown_2.11
## [37] minqa_1.2.4       purrr_0.3.4       magrittr_2.0.2    matrixcalc_1.0-5
## [41] ellipsis_0.3.2    htmltools_0.5.2   MASS_7.3-55       assertthat_0.2.1
## [45] utf8_1.2.2        stringi_1.7.6     crayon_1.5.0
```