Coefficient covariance for rotated data with random weights

Peter Hettegger

2020-11-28

This document contains R-code for analysing the coefficient covariance for two features with different (independent) sample weights. We compare coefficient covariance of non-rotated data and rotated data with random uniformly distributed sample weights. Therefore we assume a sample design with 2 groups (case, control) and sex (male, female) as second covariate (for illustration purposes).

```
set.seed(1)

library(randRotation)
library(ggplot2)
library(knitr)
library(heatmap3)

# Sample info
samp.inf <- data.frame(sex = as.factor(rep(c("male","female"), c(10,10))),
                       group = as.factor(rep(c("case", "control"), c(5,5))))
n <- nrow(samp.inf)
kable(with(samp.inf, table(sex, group)))
```

	case	control
female	5	5
male	5	5

The correlation coefficient of the error terms between feature 1 and feature 2 is assumed as 0.8. However, as it is a constant factor in the covariance of the coefficients, the value of the correlation coefficient solely changes the scales of the plots below, but not the pattern.

```
X <- model.matrix(~ 1 + sex + group, samp.inf)

rho12 <- 0.8
# assume sigma1 = sigma2 = 1
sigma12 <- rho12


# define random uniformly distributed weights
W <- matrix(runif(2*n), nrow = 2)
W
```

```
##        [,1]   [,2]   [,3]   [,4]    [,5]   [,6]   [,7]   [,8]   [,9]  [,10]
## [1,] 0.2655 0.5729 0.2017 0.9447 0.62911 0.2060 0.6870 0.7698 0.7176 0.3800
## [2,] 0.3721 0.9082 0.8984 0.6608 0.06179 0.1766 0.3841 0.4977 0.9919 0.7774
##       [,11]  [,12]  [,13]   [,14]  [,15]  [,16]  [,17]  [,18]  [,19]  [,20]
## [1,] 0.9347 0.6517 0.2672 0.01339 0.8697 0.4821 0.4935 0.8274 0.7942 0.7237
```

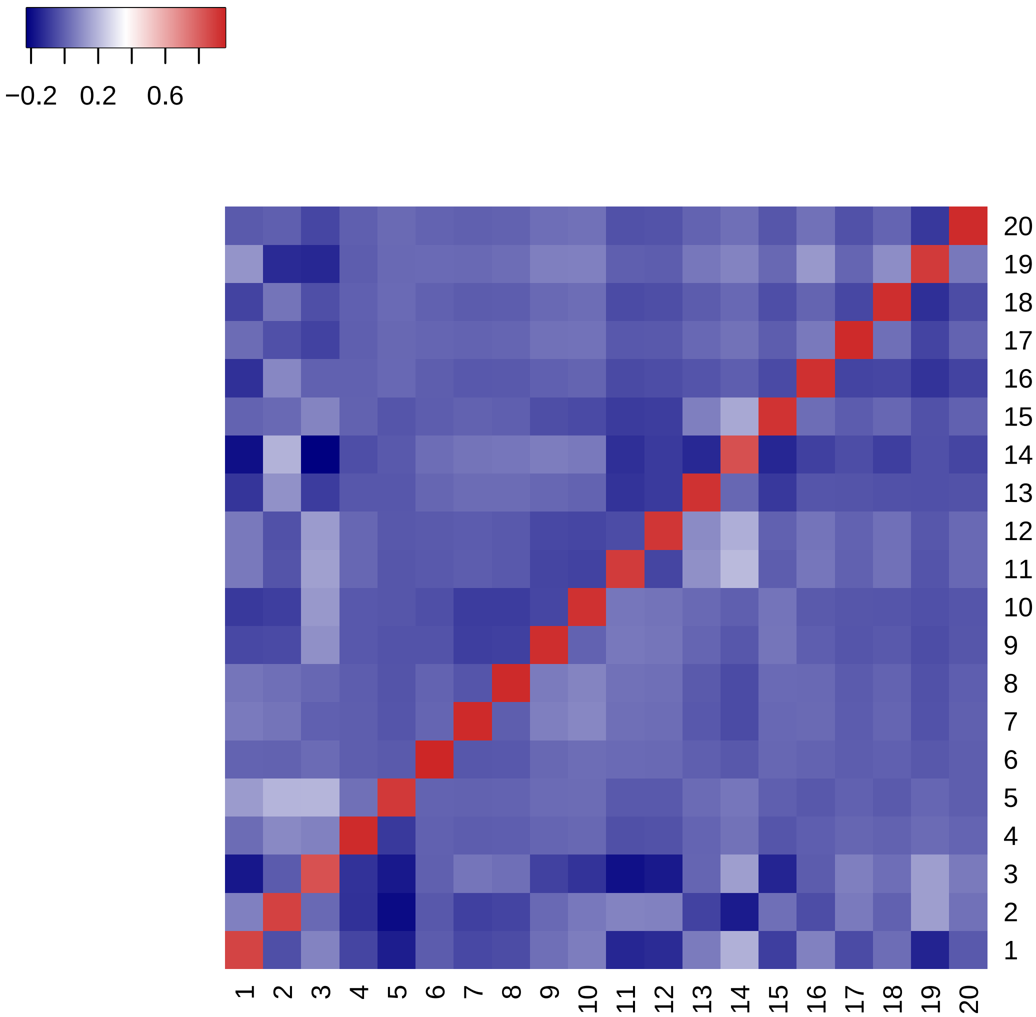
```
## [2,] 0.2121 0.1256 0.3861 0.38239 0.3403 0.5996 0.1862 0.6685 0.1079 0.4113
```

The following steps of whitening, QR-decomposition and generation of random (restricted) rotation matrices are outlined in the main manuscript.

```
# whitening of X
X1 <- sqrt(W[1,]) * X
X2 <- sqrt(W[2,]) * X


# qr decomposition

# group as "hypothesis coefficient"
# intercept and sex as "determined coefficients"
coef.d <- 1:2

Q1 <- qr.Q(qr(X1), complete = TRUE)
Xd1 <- Q1[,coef.d]
Xhe1 <- Q1[,-coef.d]

Q2 <- qr.Q(qr(X2), complete = TRUE)
Xd2 <- Q2[,coef.d]
Xhe2 <- Q2[,-coef.d]


r <- ncol(Xhe1)

E.R1R2 <- Xd1 %*% t(Xd1) %*% Xd2 %*% t(Xd2) +
    1/r * Xhe1 %*% t(Xhe2) * sum(diag(t(Xhe1)%*%Xhe2))
```

The heatmap shows the expected value `E.R1R2`

$$\mathbb{E}_R\left[\tilde{\boldsymbol{R}}^*_{r1}\,\tilde{\boldsymbol{R}}^{*T}_{r2}\right]$$



The diagonal elements of `E.R1R2` are

```
diag(E.R1R2)
```

```
##  [1] 0.8741 0.8803 0.8370 0.9394 0.9019 0.9564 0.9428 0.9438 0.9330 0.9250
## [11] 0.8972 0.9111 0.9226 0.8418 0.9192 0.9271 0.9437 0.9329 0.8989 0.9410
```

In the following, we calculate

$$\mathbb{E}_R\left[\tilde{\boldsymbol{R}}^*_{r1}\,\tilde{\boldsymbol{R}}^{*T}_{r2}\right]$$

and estimate the element wise standard deviation

$$\mathrm{sd}_R\left[\tilde{\boldsymbol{R}}^*_{r1}\,\tilde{\boldsymbol{R}}^{*T}_{r2}\right]$$

```
E.cov.beta.r <- sigma12 *
    solve(t(X1)%*%X1) %*% t(X1) %*% E.R1R2 %*% X2 %*% solve(t(X2)%*%X2)

covs <- vapply(1:100, function(i){
    R <- randorth(ncol(Xhe1))
    R1 <- Xd1 %*% t(Xd1) + Xhe1 %*% R %*% t(Xhe1)
    R2 <- Xd2 %*% t(Xd2) + Xhe2 %*% R %*% t(Xhe2)

    sigma12 * solve(t(X1)%*%X1) %*% t(X1) %*% R1 %*%
```

```r
        t(R2) %*% X2 %*% solve(t(X2)%*%X2) - E.cov.beta.r
}, matrix(1.2, 3,3))

sd.cov.beta.r <- apply(covs, 1:2, function(i)sqrt(mean(i^2)))

cov.beta <- sigma12 * solve(t(X1)%*%X1) %*% t(X1) %*% X2 %*% solve(t(X2)%*%X2)
```

Coefficient covariance for non-rotated data:

```r
kable(cov.beta)
```

	(Intercept)	sexmale	groupcontrol
(Intercept)	0.2373	-0.1641	-0.1515
sexmale	-0.1672	0.2930	0.0169
groupcontrol	-0.1484	0.0118	0.2829

Expected coefficient covariance for rotated data:

```r
kable(E.cov.beta.r)
```

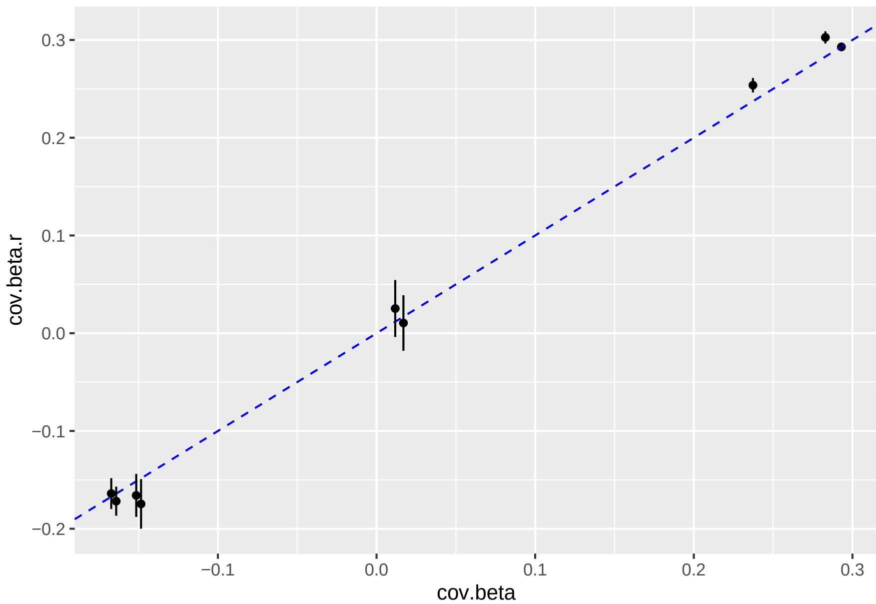
	(Intercept)	sexmale	groupcontrol
(Intercept)	0.2537	-0.1718	-0.1659
sexmale	-0.1640	0.2929	0.0105
groupcontrol	-0.1746	0.0253	0.3026

Standard deviation of coefficient covariance for rotated data:

```r
kable(sd.cov.beta.r)
```

	(Intercept)	sexmale	groupcontrol
(Intercept)	0.0074	0.0149	0.0220
sexmale	0.0158	0.0015	0.0284
groupcontrol	0.0254	0.0292	0.0063

The graphical representation of these tables is:



The given example shows, that even for random weights the dependence structure of coefficient estimates is largely retained for the assumed experimental design. The coefficient covariance can be investigated in the same manner for each individual experimental design.

Session Info

```
sessionInfo()
```

```
## R Under development (unstable) (2020-11-14 r79432)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19041)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=German_Austria.1252  LC_CTYPE=German_Austria.1252
## [3] LC_MONETARY=German_Austria.1252 LC_NUMERIC=C
## [5] LC_TIME=German_Austria.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] heatmap3_1.1.7     knitr_1.30         ggplot2_3.3.2      randRotation_1.3.4
##
## loaded via a namespace (and not attached):
##  [1] xml2_1.3.2         magrittr_1.5       munsell_0.5.0      colorspace_2.0-0
##  [5] R6_2.5.0           rlang_0.4.8        highr_0.8          fastcluster_1.1.25
##  [9] stringr_1.4.0      tools_4.1.0        rbibutils_1.4      grid_4.1.0
## [13] gtable_0.3.0       xfun_0.19          withr_2.3.0        ellipsis_0.3.1
## [17] htmltools_0.5.0    yaml_2.2.1         digest_0.6.27      tibble_3.0.4
## [21] lifecycle_0.2.0    crayon_1.3.4       farver_2.0.3       vctrs_0.3.4
## [25] Rdpack_2.1         gbRd_0.4-11        glue_1.4.2         evaluate_0.14
## [29] rmarkdown_2.5      labeling_0.4.2     stringi_1.5.3      pillar_1.4.6
## [33] compiler_4.1.0     scales_1.1.1       pkgconfig_2.0.3
```