# Coefficient covariance for rotated data with identical weights

Peter Hettegger

2020-11-28

This document contains R-code for analysing the coefficient covariance for two features with sample weights equal 1. We compare coefficient covariance of non-rotated data and rotated data. Therefore we assume a sample design with `age` and `group` (case, control) as covariates. `group` is the hypothesis coefficient.

```
set.seed(1)

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

# Sample info
samp.inf <- data.frame(age = as.integer(runif(20)*30),
                       group = as.factor(rep(c("case", "control"), c(10,10))))
n <- nrow(samp.inf)
kable(head(samp.inf))
```

| age | group |
|---:|---|
| 7 | case |
| 11 | case |
| 17 | case |
| 27 | case |
| 6 | case |
| 26 | case |

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 + age + group, samp.inf)

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


# define random uniformly distributed weights
W <- matrix(runif(10), nrow = 2, ncol = n, byrow = TRUE)
W
```

```
##        [,1]   [,2]   [,3]   [,4]   [,5]   [,6]    [,7]   [,8]   [,9]  [,10]
## [1,] 0.9347 0.2121 0.6517 0.1256 0.2672 0.3861 0.01339 0.3824 0.8697 0.3403
```

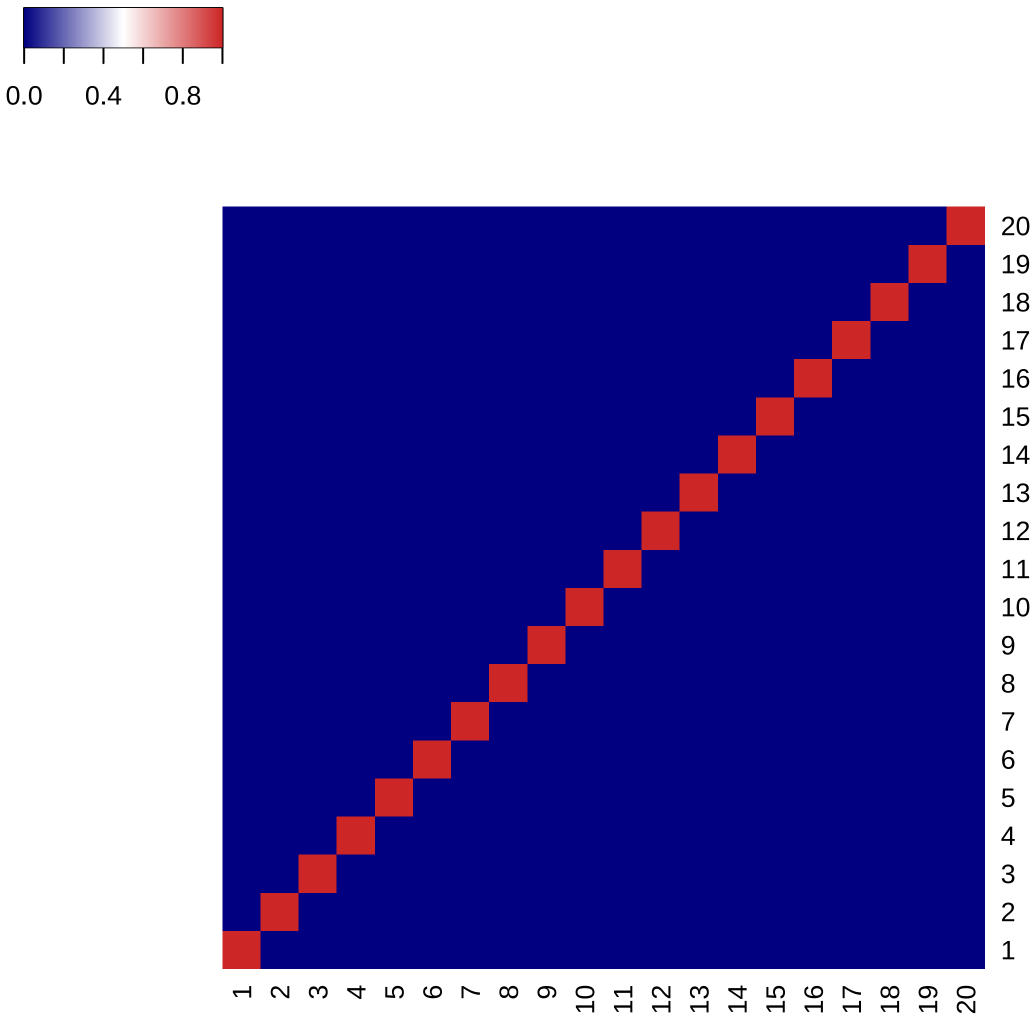
```
## [2,] 0.9347 0.2121 0.6517 0.1256 0.2672 0.3861 0.01339 0.3824 0.8697 0.3403
##       [,11]  [,12]  [,13]  [,14]  [,15]  [,16]   [,17]  [,18]  [,19]  [,20]
## [1,] 0.9347 0.2121 0.6517 0.1256 0.2672 0.3861 0.01339 0.3824 0.8697 0.3403
## [2,] 0.9347 0.2121 0.6517 0.1256 0.2672 0.3861 0.01339 0.3824 0.8697 0.3403
```

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

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


# qr decomposition

# group as "hypothesis coefficient"
# intercept and age 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] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

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 %*%
        t(R2) %*% X2 %*% solve(t(X2)%*%X2) - E.cov.beta.r
```

```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) | age | groupcontrol |
|---|---|---|---|
| (Intercept) | 0.5223 | -0.0236 | -0.1756 |
| age | -0.0236 | 0.0017 | -0.0011 |
| groupcontrol | -0.1756 | -0.0011 | 0.3832 |

Expected coefficient covariance for rotated data:

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

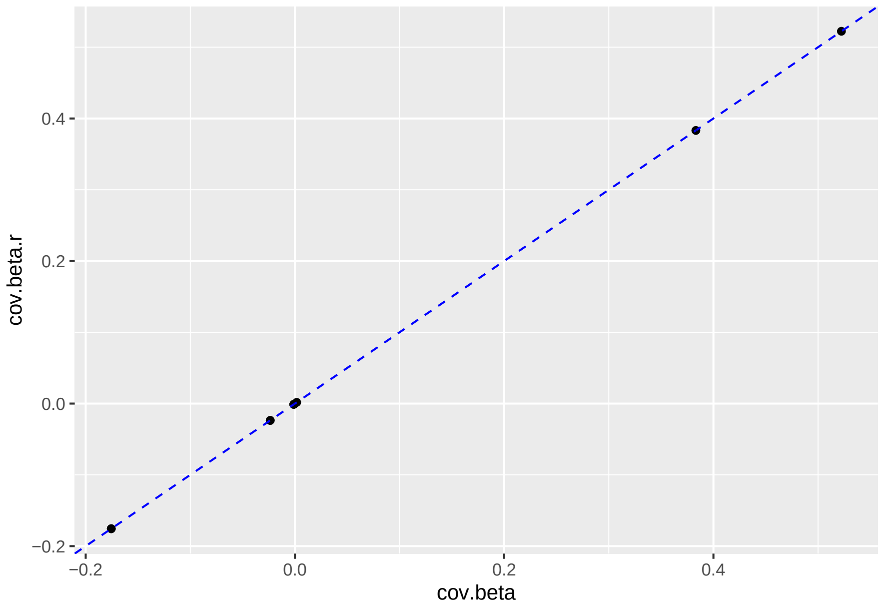
| | (Intercept) | age | groupcontrol |
|---|---|---|---|
| (Intercept) | 0.5223 | -0.0236 | -0.1756 |
| age | -0.0236 | 0.0017 | -0.0011 |
| groupcontrol | -0.1756 | -0.0011 | 0.3832 |

Standard deviation of coefficient covariance for rotated data:

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

| | (Intercept) | age | groupcontrol |
|---|---|---|---|
| (Intercept) | 0 | 0 | 0 |
| age | 0 | 0 | 0 |
| groupcontrol | 0 | 0 | 0 |

The graphical representation of these tables is:




The given example shows, that for identical weights, the dependence structure of coefficient estimates is exactly retained with random rotation for the assumed 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
```