Introduction to RBM package

Dongmei Li

October 30, 2025

Clinical and Translational Science Institute, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642-0708

Contents

1 Overview

This document provides an introduction to the `RBM` package. The `RBM` package executes the resampling-based empirical Bayes approach using either permutation or bootstrap tests based on moderated t-statistics through the following steps.

- Firstly, the RBM package computes the moderated t-statistics based on the observed data set for each feature using the lmFit and eBayes function.
- Secondly, the original data are permuted or bootstrapped in a way that matches the null hypothesis to generate permuted or bootstrapped resamples, and the reference distribution is constructed using the resampled moderated t-statistics calculated from permutation or bootstrap resamples.
- Finally, the p-values from permutation or bootstrap tests are calculated based on the proportion of the permuted or bootstrapped moderated t-statistics that are as extreme as, or more extreme than, the observed moderated t-statistics.

Additional detailed information regarding resampling-based empirical Bayes approach can be found elsewhere (Li et al., 2013).

2 Getting started

The RBM package can be installed and loaded through the following R code.
Install the RBM package with:

```
> if (!requireNamespace("BiocManager", quietly=TRUE))
+     install.packages("BiocManager")
> BiocManager::install("RBM")
```

Load the RBM package with:

```
> library(RBM)
```

3 RBM_T and RBM_F functions

There are two functions in the RBM package: RBM_T and RBM_F. Both functions require input data in the matrix format with rows denoting features and columns denoting samples. RBM_T is used for two-group comparisons such as study designs with a treatment group and a control group. RBM_F can be used for more complex study designs such as more than two groups or time-course studies. Both functions need a vector for group notation, i.e., "1" denotes the treatment group and "0" denotes the control group. For the RBM_F function, a contrast vector need to be provided by users to perform pairwise comparisons between groups. For example, if the design has three groups (0, 1, 2), the aContrast parameter will be a vector such as ("X1-X0", "X2-X1", "X2-X0") to denote all pairwise comparisons. Users just need to add an extra "X" before the group labels to do the contrasts.

- Examples using the RBM_T function: normdata simulates a standardized gene expression data and unifdata simulates a methylation microarray data. The p-values from the RBM_T function could be further adjusted using the p.adjust function in the stats package through the Bejamini-Hochberg method.

```
> library(RBM)
> normdata <- matrix(rnorm(1000*6, 0, 1),1000,6)
> mydesign <- c(0,0,0,1,1,1)
> myresult <- RBM_T(normdata,mydesign,100,0.05)
> summary(myresult)

              Length Class  Mode
ordfit_t      1000   -none- numeric
ordfit_pvalue 1000   -none- numeric
ordfit_beta0  1000   -none- numeric
ordfit_beta1  1000   -none- numeric
permutation_p 1000   -none- numeric
bootstrap_p   1000   -none- numeric

> sum(myresult$permutation_p<=0.05)
```

```
[1] 30

> which(myresult$permutation_p<=0.05)

 [1]   7  16  19  81 101 153 248 277 294 326 396 419 433 436 488 503 555 563 646
[20] 679 690 719 797 813 850 863 888 949 978 989

> sum(myresult$bootstrap_p<=0.05)

[1] 7

> which(myresult$bootstrap_p<=0.05)

[1]  81 153 199 237 436 649 978

> permutation_adjp <- p.adjust(myresult$permutation_p, "BH")
> sum(permutation_adjp<=0.05)

[1] 4

> bootstrap_adjp <- p.adjust(myresult$bootstrap_p, "BH")
> sum(bootstrap_adjp<=0.05)

[1] 0

> unifdata <- matrix(runif(1000*7,0.10, 0.95), 1000, 7)
> mydesign2 <- c(0,0,0, 1,1,1,1)
> myresult2 <- RBM_T(unifdata,mydesign2,100,0.05)
> sum(myresult2$permutatioin_p<=0.05)

[1] 0

> sum(myresult2$bootstrap_p<=0.05)

[1] 14

> which(myresult2$bootstrap_p<=0.05)

 [1] 230 266 272 321 390 426 457 560 584 647 819 866 909 990

> bootstrap2_adjp <- p.adjust(myresult2$bootstrap_p, "BH")
> sum(bootstrap2_adjp<=0.05)

[1] 0
```

- Examples using the `RBM_F` function: normdata_F simulates a standardized gene expression data and unifdata_F simulates a methylation microarray data. In both examples, we were interested in pairwise comparisons.

```
> normdata_F <- matrix(rnorm(1000*9,0,2), 1000, 9)
> mydesign_F <- c(0, 0, 0, 1, 1, 1, 2, 2, 2)
> aContrast <- c("X1-X0", "X2-X1", "X2-X0")
> myresult_F <- RBM_F(normdata_F, mydesign_F, aContrast, 100, 0.05)
> summary(myresult_F)

             Length Class  Mode
ordfit_t     3000   -none- numeric
ordfit_pvalue 3000   -none- numeric
ordfit_beta1  3000   -none- numeric
permutation_p 3000   -none- numeric
bootstrap_p   3000   -none- numeric

> sum(myresult_F$permutation_p[, 1]<=0.05)

[1] 53

> sum(myresult_F$permutation_p[, 2]<=0.05)

[1] 57

> sum(myresult_F$permutation_p[, 3]<=0.05)

[1] 55

> which(myresult_F$permutation_p[, 1]<=0.05)

 [1]  20  33  37  60  83  90  97 123 125 154 168 176 205 217 218 238 278 289 314
[20] 315 318 335 365 377 426 436 460 479 492 519 541 554 571 584 621 643 653 677
[39] 687 709 718 739 741 769 773 784 820 840 853 859 891 907 971

> which(myresult_F$permutation_p[, 2]<=0.05)

 [1]  20  31  33  37  56  60  83  88  90  97 123 154 161 168 176 218 221 238 278
[20] 286 289 298 314 315 318 335 347 365 372 377 426 440 460 492 507 519 541 554
[39] 571 584 621 687 705 709 721 739 769 773 833 840 859 868 904 932 936 947 971

> which(myresult_F$permutation_p[, 3]<=0.05)

 [1]  20  33  37  41  56  60  97 123 161 168 176 211 217 218 221 234 238 278 289
[20] 315 318 335 347 365 377 426 436 460 492 507 519 541 554 571 584 621 643 677
[39] 687 709 739 741 769 773 784 820 833 840 859 891 904 907 932 947 971

> con1_adjp <- p.adjust(myresult_F$permutation_p[, 1], "BH")
> sum(con1_adjp<=0.05/3)

[1] 6
```

```
> con2_adjp <- p.adjust(myresult_F$permutation_p[, 2], "BH")
> sum(con2_adjp<=0.05/3)

[1] 13

> con3_adjp <- p.adjust(myresult_F$permutation_p[, 3], "BH")
> sum(con3_adjp<=0.05/3)

[1] 8

> which(con2_adjp<=0.05/3)

 [1] 123 168 289 315 335 377 460 492 541 554 584 769 840

> which(con3_adjp<=0.05/3)

[1] 123 168 315 335 426 519 541 773

> unifdata_F <- matrix(runif(1000*18, 0.15, 0.98), 1000, 18)
> mydesign2_F <- c(rep(0, 6), rep(1, 6), rep(2, 6))
> aContrast <- c("X1-X0", "X2-X1", "X2-X0")
> myresult2_F <- RBM_F(unifdata_F, mydesign2_F, aContrast, 100, 0.05)
> summary(myresult2_F)

             Length Class  Mode
ordfit_t     3000   -none- numeric
ordfit_pvalue 3000   -none- numeric
ordfit_beta1  3000   -none- numeric
permutation_p 3000   -none- numeric
bootstrap_p   3000   -none- numeric

> sum(myresult2_F$bootstrap_p[, 1]<=0.05)

[1] 58

> sum(myresult2_F$bootstrap_p[, 2]<=0.05)

[1] 44

> sum(myresult2_F$bootstrap_p[, 3]<=0.05)

[1] 49

> which(myresult2_F$bootstrap_p[, 1]<=0.05)

 [1]  12  27  49  58  68  75 131 173 182 187 191 207 211 266 276 280 314 340 377
[20] 396 399 417 423 426 467 482 484 491 497 498 500 532 553 572 579 592 600 619
[39] 624 639 670 699 731 737 803 806 815 817 825 832 834 841 864 888 953 972 978
[58] 980
```

```
> which(myresult2_F$bootstrap_p[, 2]<=0.05)

 [1]  27  68  75 131 187 207 211 266 314 340 377 394 399 417 467 482 491 497 498
[20] 500 532 553 579 592 619 639 670 690 699 731 737 753 806 815 817 825 832 834
[39] 864 932 953 972 978 980

> which(myresult2_F$bootstrap_p[, 3]<=0.05)

 [1]  12  27  58  92 111 173 187 207 211 266 276 278 314 340 377 399 417 467 491
[20] 497 498 500 528 532 579 592 619 624 627 635 639 699 731 737 803 806 815 817
[39] 825 832 834 841 845 873 888 932 972 978 980

> con21_adjp <- p.adjust(myresult2_F$bootstrap_p[, 1], "BH")
> sum(con21_adjp<=0.05/3)

[1] 8

> con22_adjp <- p.adjust(myresult2_F$bootstrap_p[, 2], "BH")
> sum(con22_adjp<=0.05/3)

[1] 6

> con23_adjp <- p.adjust(myresult2_F$bootstrap_p[, 3], "BH")
> sum(con23_adjp<=0.05/3)

[1] 4
```

4 Ovarian cancer methylation example using the RBM_T function

Two-group comparisons are the most common contrast in biological and biomedical field. The ovarian cancer methylation example is used to illustrate the application of RBM_T in identifying differentially methylated loci. The ovarian cancer methylation example is taken from the gemone-wide DNA methylation profiling of United Kingdom Ovarian Cancer Population Study (UKOPS). This study used Illumina Infinium 27k Human DNA methylation Beadchip v1.2 to obtain DNA methylation profiles on over 27,000 CpGs in whole blood cells from 266 ovarian cancer women and 274 age-matched healthy controls. The data are downloaded from the NCBI GEO website with access number GSE19711. For illutration purpose, we chose the first 1000 loci in 8 randomly selected women with 4 ovariance cancer cases (pre-treatment) and 4 healthy controls. The following codes show the process of generating significant differential DNA methylation loci using the RBM_T function and presenting the results for further validation and investigations.

```
> system.file("data", package = "RBM")

[1] "/tmp/Rtmpu193Se/Rinst1dc8bb46f9f345/RBM/data"

> data(ovarian_cancer_methylation)
> summary(ovarian_cancer_methylation)
```

```
       IlmnID         Beta          exmdata2[, 2]     exmdata3[, 2]
 cg00000292:  1   Min.   :0.01058   Min.   :0.01187   Min.   :0.009103
 cg00002426:  1   1st Qu.:0.04111   1st Qu.:0.04407   1st Qu.:0.041543
 cg00003994:  1   Median :0.08284   Median :0.09531   Median :0.087042
 cg00005847:  1   Mean   :0.27397   Mean   :0.28872   Mean   :0.283729
 cg00006414:  1   3rd Qu.:0.52135   3rd Qu.:0.59031   3rd Qu.:0.558575
 cg00007981:  1   Max.   :0.97069   Max.   :0.96937   Max.   :0.970155
 (Other)   :994                     NA's   :4
 exmdata4[, 2]     exmdata5[, 2]     exmdata6[, 2]     exmdata7[, 2]
 Min.   :0.01019   Min.   :0.01108   Min.   :0.01937   Min.   :0.01278
 1st Qu.:0.04092   1st Qu.:0.04059   1st Qu.:0.05060   1st Qu.:0.04260
 Median :0.09042   Median :0.08527   Median :0.09502   Median :0.09362
 Mean   :0.28508   Mean   :0.28482   Mean   :0.27348   Mean   :0.27563
 3rd Qu.:0.57502   3rd Qu.:0.57300   3rd Qu.:0.52099   3rd Qu.:0.52240
 Max.   :0.96658   Max.   :0.97516   Max.   :0.96681   Max.   :0.95974
                   NA's   :1
 exmdata8[, 2]
 Min.   :0.01357
 1st Qu.:0.04387
 Median :0.09282
 Mean   :0.28679
 3rd Qu.:0.57217
 Max.   :0.96268
```

```
> ovarian_cancer_data <- ovarian_cancer_methylation[, -1]
> label <- c(1, 1, 0, 0, 1, 1, 0, 0)
> diff_results <- RBM_T(aData=ovarian_cancer_data, vec_trt=label, repetition=100, alpha=0.05)
> summary(diff_results)

             Length Class  Mode
ordfit_t     1000   -none- numeric
ordfit_pvalue 1000   -none- numeric
ordfit_beta0  1000   -none- numeric
ordfit_beta1  1000   -none- numeric
permutation_p 1000   -none- numeric
bootstrap_p   1000   -none- numeric

> sum(diff_results$ordfit_pvalue<=0.05)

[1] 47

> sum(diff_results$permutation_p<=0.05)

[1] 55

> sum(diff_results$bootstrap_p<=0.05)
```

```
[1] 82
```

```
> ordfit_adjp <- p.adjust(diff_results$ordfit_pvalue, "BH")
> sum(ordfit_adjp<=0.05)
```

```
[1] 0
```

```
> perm_adjp <- p.adjust(diff_results$permutation_p, "BH")
> sum(perm_adjp<=0.05)
```

```
[1] 1
```

```
> boot_adjp <- p.adjust(diff_results$bootstrap_p, "BH")
> sum(boot_adjp<=0.05)
```

```
[1] 14
```

```
> diff_list_perm <- which(perm_adjp<=0.05)
> diff_list_boot <- which(boot_adjp<=0.05)
> sig_results_perm <- cbind(ovarian_cancer_methylation[diff_list_perm, ], diff_results$ordfit_t[
> print(sig_results_perm)
```

```
       IlmnID      Beta exmdata2[, 2] exmdata3[, 2] exmdata4[, 2] exmdata5[, 2]
19 cg00016968 0.8062848            NA     0.8144082     0.8362318     0.8083138
   exmdata6[, 2] exmdata7[, 2] exmdata8[, 2]
19     0.7330644     0.8296834      0.849178
   diff_results$ordfit_t[diff_list_perm]
19                             -2.547097
   diff_results$permutation_p[diff_list_perm]
19                                      0
```

```
> sig_results_boot <- cbind(ovarian_cancer_methylation[diff_list_boot, ], diff_results$ordfit_t[
> print(sig_results_boot)
```

```
        IlmnID       Beta exmdata2[, 2] exmdata3[, 2] exmdata4[, 2]
50  cg00044245 0.07833433    0.08098538    0.09683810    0.09055518
95  cg00081975 0.03633894    0.04975194    0.06024723    0.05598723
131 cg00121904 0.15449580    0.17949750    0.23608110    0.24354150
146 cg00134539 0.61101320    0.53321780    0.45999340    0.46787420
258 cg00234616 0.06639040    0.14705640    0.18254770    0.19942150
259 cg00234961 0.04192170    0.04321576    0.05707140    0.05327565
278 cg00256281 0.01924708    0.03364037    0.01946214    0.02172512
346 cg00331237 0.05972383            NA    0.08204769    0.08345662
482 cg00468146 0.11144740    0.15416650    0.19827990    0.18517240
632 cg00615377 0.11265030    0.16140570    0.19404450    0.17468600
677 cg00651216 0.06825629    0.12529090    0.14409190    0.13907250
736 cg00706536 0.43475180    0.48246400    0.52171670    0.54656730
```

```
887 cg00862290 0.43640520    0.54047160    0.60786800    0.56325950
979 cg00945507 0.13432250    0.23854600    0.34749760    0.28903340
    exmdata5[, 2] exmdata6[, 2] exmdata7[, 2] exmdata8[, 2]
50     0.08555133    0.08750076    0.08391400    0.10606830
95     0.04561792    0.05115624    0.06068253    0.06168212
131    0.17352980    0.12564280    0.18193170    0.20847670
146    0.67191510    0.63137380    0.47929610    0.45428300
258    0.10620550    0.11668540    0.12630260    0.19163650
259    0.04030003    0.03996053    0.05086962    0.05445672
278    0.02644337    0.03560722    0.01365959    0.02060407
346    0.05372019    0.06241126    0.06955040    0.09140985
482    0.12285820    0.13271110    0.14196260    0.22159420
632    0.12573100    0.14483660    0.16338240    0.20130510
677    0.07669587    0.09597587    0.11690440    0.15194540
736    0.46315280    0.47922440    0.47347480    0.50734820
887    0.50259740    0.40111730    0.56646700    0.54552980
979    0.11848510    0.16653850    0.30718420    0.26624740
    diff_results$ordfit_t[diff_list_boot]
50                              -1.825937
95                              -2.654324
131                             -3.562745
146                              5.636263
258                             -3.168405
259                             -2.833203
278                              1.777836
346                             -3.328798
482                             -3.318963
632                             -3.722206
677                             -3.457874
736                             -2.822475
887                             -3.368752
979                             -4.968792
    diff_results$bootstrap_p[diff_list_boot]
50                                        0
95                                        0
131                                       0
146                                       0
258                                       0
259                                       0
278                                       0
346                                       0
482                                       0
632                                       0
677                                       0
736                                       0
```

887	0
979	0