Analysis of BMI GIANT GWAS data - Scott approach

Simina Boca, Jeff Leek

2017-06-09

Load the relevant libraries:

```
library(readr)
```

```
## Warning: package 'readr' was built under R version 3.3.3
```

```
library(dplyr)
library(fdrtool)
```

```
## Warning: package 'fdrtool' was built under R version 3.3.2
```

```
library(betareg)
```

```
## Warning: package 'betareg' was built under R version 3.3.3
```

```
library(splines)
library(Hmisc)
```

```
## Warning: package 'Hmisc' was built under R version 3.3.3
```

```
## Warning: package 'survival' was built under R version 3.3.3
```

```
## Warning: package 'Formula' was built under R version 3.3.2
```

```
## Warning: package 'ggplot2' was built under R version 3.3.3
```

```
library(ggplot2)
library(reshape2)
library(FDRreg)
```

```
## Warning: package 'fda' was built under R version 3.3.3
```

```
## Warning: package 'BayesLogit' was built under R version 3.3.2
```

```
## Warning: package 'mvtnorm' was built under R version 3.3.2
```

Load the .RData file with the BMI GIANT GWAS meta-analysis data:

```
load("BMI_GIANT_GWAS.RData")

tot <- BMI_GIANT_GWAS
```

Estimate fraction of true null hypotheses in a regression framework using the Scott approach

Create the design matrix, using natural cubic splines with 5 degrees of freedom to model `N` and 3 discrete categories for the MAFs:

```
X <- model.matrix(~ splines::ns(N,5) + Freq_MAF_Int_Hapmap, data = tot)[,-1]
dim(X)
```

```
## [1] 2500573       7
```

```
head(X)
```

```
##   splines::ns(N, 5)1 splines::ns(N, 5)2 splines::ns(N, 5)3
## 1       4.414107e-01       5.538398e-01      -0.0017421409
## 2       0.000000e+00       3.954615e-10      -0.1655612193
## 3       3.884106e-05       9.880678e-01       0.0099138318
## 4       3.209714e-04       9.891039e-01       0.0088153851
## 5       9.327150e-02       9.061998e-01       0.0002901038
## 6       6.724476e-04       9.894590e-01       0.0082264435
##   splines::ns(N, 5)4 splines::ns(N, 5)5 Freq_MAF_Int_Hapmap[0.127,0.302)
## 1       0.0034871548      -0.0017450139                                 0
## 2       0.3336072837       0.8319539352                                 0
## 3       0.0039611703      -0.0019816905                                 1
## 4       0.0035221381      -0.0017623939                                 1
## 5       0.0002264943      -0.0001133405                                 1
## 6       0.0032867973      -0.0016447160                                 0
##   Freq_MAF_Int_Hapmap[0.302,0.500]
## 1                                1
## 2                                1
## 3                                0
## 4                                0
## 5                                0
## 6                                1
```

Run code to estimate the fraction of true null hypotheses within a regression framework with the design matrix specified above:

```
##first get z-scores, which are needed for the Scott approach
zScores <- tot$b/tot$se

range(zScores)
```

```
## [1] -22.41379  26.96667
```

```
##run FDRreg
fdr <- FDRreg(zScores, X,
              nulltype = 'theoretical',
              control=list(lambda=1))

##get prior probabilities
pi0EstScott <- 1-fdr$priorprob

##get FDR
FDRScott_theo <- fdr$FDR
```

Save results:

```
save(list=c("pi0EstScott","FDRScott_theo"), file="BMI_GIANT_GWAS_results_Scott_theoretical.RData")
```

Session Information

```
devtools::session_info()
```

```
## Session info -------------------------------------------------------------
```

```
##  setting  value
##  version  R version 3.3.1 (2016-06-21)
##  system   x86_64, mingw32
```

```
##  ui       RTerm
##  language (EN)
##  collate  English_United States.1252
##  tz       America/New_York
##  date     2017-06-09
```

```
## Packages -----------------------------------------------------------------
```

```
##  package      * version date       source
##  acepack        1.4.1   2016-10-29 CRAN (R 3.3.3)
##  assertthat     0.1     2013-12-06 CRAN (R 3.3.1)
##  backports      1.0.4   2016-10-24 CRAN (R 3.3.1)
##  base64enc      0.1-3   2015-07-28 CRAN (R 3.3.2)
##  BayesLogit   * 0.6     2016-10-20 CRAN (R 3.3.2)
##  betareg      * 3.1-0   2016-08-06 CRAN (R 3.3.3)
##  BiocStyle    * 2.0.3   2016-08-04 Bioconductor
##  checkmate      1.8.2   2016-11-02 CRAN (R 3.3.3)
##  cluster        2.0.4   2016-04-18 CRAN (R 3.3.1)
##  codetools      0.2-14  2015-07-15 CRAN (R 3.3.1)
##  colorspace     1.2-6   2015-03-11 CRAN (R 3.3.1)
##  data.table     1.10.4  2017-02-01 CRAN (R 3.3.2)
##  DBI            0.4-1   2016-05-08 CRAN (R 3.3.1)
##  devtools       1.12.0  2016-06-24 CRAN (R 3.3.3)
##  digest         0.6.9   2016-01-08 CRAN (R 3.3.1)
##  dplyr        * 0.4.3   2015-09-01 CRAN (R 3.3.1)
##  evaluate       0.10    2016-10-11 CRAN (R 3.3.1)
##  fda          * 2.4.4   2014-12-16 CRAN (R 3.3.3)
##  FDRreg       * 0.2-1   2017-05-03 Github (jgscott/FDRreg@8025d1a)
##  fdrtool      * 1.2.15  2015-07-08 CRAN (R 3.3.2)
##  flexmix        2.3-14  2017-04-28 CRAN (R 3.3.3)
##  foreign        0.8-66  2015-08-19 CRAN (R 3.3.1)
##  Formula      * 1.2-1   2015-04-07 CRAN (R 3.3.2)
##  ggdendro       0.1-20  2016-04-27 CRAN (R 3.3.3)
##  ggplot2      * 2.2.1   2016-12-30 CRAN (R 3.3.3)
##  gridExtra      2.2.1   2016-02-29 CRAN (R 3.3.1)
##  gtable         0.2.0   2016-02-26 CRAN (R 3.3.1)
##  Hmisc        * 4.0-3   2017-05-02 CRAN (R 3.3.3)
##  hms            0.3     2016-11-22 CRAN (R 3.3.3)
##  htmlTable      1.9     2017-01-26 CRAN (R 3.3.3)
##  htmltools      0.3.5   2016-03-21 CRAN (R 3.3.1)
##  htmlwidgets    0.8     2016-11-09 CRAN (R 3.3.3)
##  knitr          1.15.1  2016-11-22 CRAN (R 3.3.1)
##  lattice      * 0.20-33 2015-07-14 CRAN (R 3.3.1)
##  latticeExtra   0.6-28  2016-02-09 CRAN (R 3.3.3)
##  lazyeval       0.2.0   2016-06-12 CRAN (R 3.3.1)
##  lmtest         0.9-35  2017-02-11 CRAN (R 3.3.3)
##  magrittr       1.5     2014-11-22 CRAN (R 3.3.1)
##  MASS           7.3-45  2016-04-21 CRAN (R 3.3.1)
##  Matrix       * 1.2-6   2016-05-02 CRAN (R 3.3.1)
##  memoise        1.0.0   2016-01-29 CRAN (R 3.3.1)
##  modeltools     0.2-21  2013-09-02 CRAN (R 3.3.2)
##  mosaic         0.14.4  2016-07-29 CRAN (R 3.3.3)
##  mosaicData     0.14.0  2016-06-17 CRAN (R 3.3.3)
##  munsell        0.4.3   2016-02-13 CRAN (R 3.3.1)
##  mvtnorm      * 1.0-6   2017-03-02 CRAN (R 3.3.2)
```

```
##  nnet           7.3-12  2016-02-02 CRAN (R 3.3.1)
##  plyr           1.8.4   2016-06-08 CRAN (R 3.3.1)
##  R6             2.1.2   2016-01-26 CRAN (R 3.3.1)
##  RColorBrewer   1.1-2   2014-12-07 CRAN (R 3.3.0)
##  Rcpp           0.12.10 2017-03-19 CRAN (R 3.3.3)
##  readr        * 1.1.0   2017-03-22 CRAN (R 3.3.3)
##  reshape2     * 1.4.1   2014-12-06 CRAN (R 3.3.1)
##  rmarkdown      1.2     2016-11-21 CRAN (R 3.3.1)
##  rpart          4.1-10  2015-06-29 CRAN (R 3.3.1)
##  rprojroot      1.1     2016-10-29 CRAN (R 3.3.1)
##  sandwich       2.3-4   2015-09-24 CRAN (R 3.3.3)
##  scales         0.4.1   2016-11-09 CRAN (R 3.3.3)
##  stringi        1.1.1   2016-05-27 CRAN (R 3.3.0)
##  stringr        1.0.0   2015-04-30 CRAN (R 3.3.1)
##  survival     * 2.41-3  2017-04-04 CRAN (R 3.3.3)
##  tibble         1.2     2016-08-26 CRAN (R 3.3.2)
##  tidyr          0.5.1   2016-06-14 CRAN (R 3.3.1)
##  withr          1.0.2   2016-06-20 CRAN (R 3.3.1)
##  yaml           2.1.13  2014-06-12 CRAN (R 3.3.1)
##  zoo            1.7-14  2016-12-16 CRAN (R 3.3.2)
```