

# HW6: Power-FDR curves of the regularization path of LASSO

Dec 4, 2023

## HW6: Power and FDR over the LASSO regularization path.

Due: December 5, 10pm in D2L.

The following code simulates data for 3,000 observations and 1,000 predictors. Ten of the 1,000 predictors have non-zero effects. These 10 predictors explain 25% of the variance of the response (y).

```
# Parameters
n=3000
p=1000
RSq=.25
q=10 # <del>number of predictors with non-zero effects</del>

# One Monte Carlo replicate
X=matrix(nrow=n,ncol=p,rnorm(n*p))
X=scale(X,center=TRUE,scale=FALSE)
causal=sample(1:p,size=q,replace=FALSE)

signal=X[,causal]*%*%rexp(q)
noise=rnorm(n,sd=sqrt(var(signal)*(1-RSq)/RSq))
y=signal+noise
# end of Monte Carlo replicate
```

The following code fits LASSO over a pre-determined grid of values of the regularization parameter (lambda).

```
library(glmnet)

## Loading required package: Matrix
## Loaded glmnet 4.1-8

lambda=exp(seq(from=log(2.85),to=log(0.0005),length=100))
fm=glmnet(y=y,x=X,lambda=lambda)
```

The estimated effects are returned in `fm$beta` (I recommend you convert it to matrix `B=as.matrix(fm$beta)`). Each column of the matrix gives the estimated effects of each predictor for a value of lambda. For any value of lambda, the discovery set (i.e., the set of active predictors) can be found using `which(B[,j]!=0)`. Of these discoveries some are true (i.e., are predictors that have non-zero effect in the simulation) and others are false.

For one Monte Carlo simulation you can calculate the proportion of false discoveries over the grid of values of lambda, as well as the proportion of effects with non-zero effect that are part of the discovery set. Averaging these results over several Monte Carlo replicates produces estimates of the power and FDR for each value of lambda.

**Goal:** To estimate power-fdr curves over the regularization path (values of lambda) of the LASSO regression. To create these curves use at least 50 Monte Carlo Simulations. For each Monte Carlo simulation, fit LASSO using the pre-specified grid of values of lambda.

**Hint:** You can create a matrix `FDP=matrix(nrow=length(lambda),ncol=nRep,NA)` and a similar matrix for power (say, `PWR`), and, store in columns the results of each Monte Carlo Simulation. To

get your estimates of FDR by value of lambda you can use `fdr=rowMeans(FDP,na.rm=TRUE)` and `power=rowMeans(PWR,na.rm=TRUE)`. Then, you can use `plot(power~fdr,type='o')` to display your results.

**Question 1:** Provide plots showing power-fdr curves for the following scenarios (1 plot per scenario):

- $RSq=0.25, n=3000$
- $RSq=0.4, n=3000$
- $RSq=0.25, n=5000$
- $RSq=0.4, n=5000$

```
library(glmnet)
SCEN=expand.grid(RSq=c(.25,.4),n=c(3000,5000),FDRPower50=NA,PowerFDR10=NA)
nRep=100

par(mfrow=c(2,2))

for(scen in 1:nrow(SCEN)){
  n=SCEN$n[scen]
  RSq=SCEN$RSq[scen]
  p=1000
  q=10

  # Objects to store results of each monte carlo simulation
  FDP=matrix(nrow=100,ncol=nRep)
  PWR=FDP

  for(mcRep in 1:nRep){
    # Simulation
    X=matrix(nrow=n,ncol=p,rnorm(n*p))
    X=scale(X,center=TRUE,scale=FALSE)
    causal=sample(1:p,size=q,replace=FALSE)

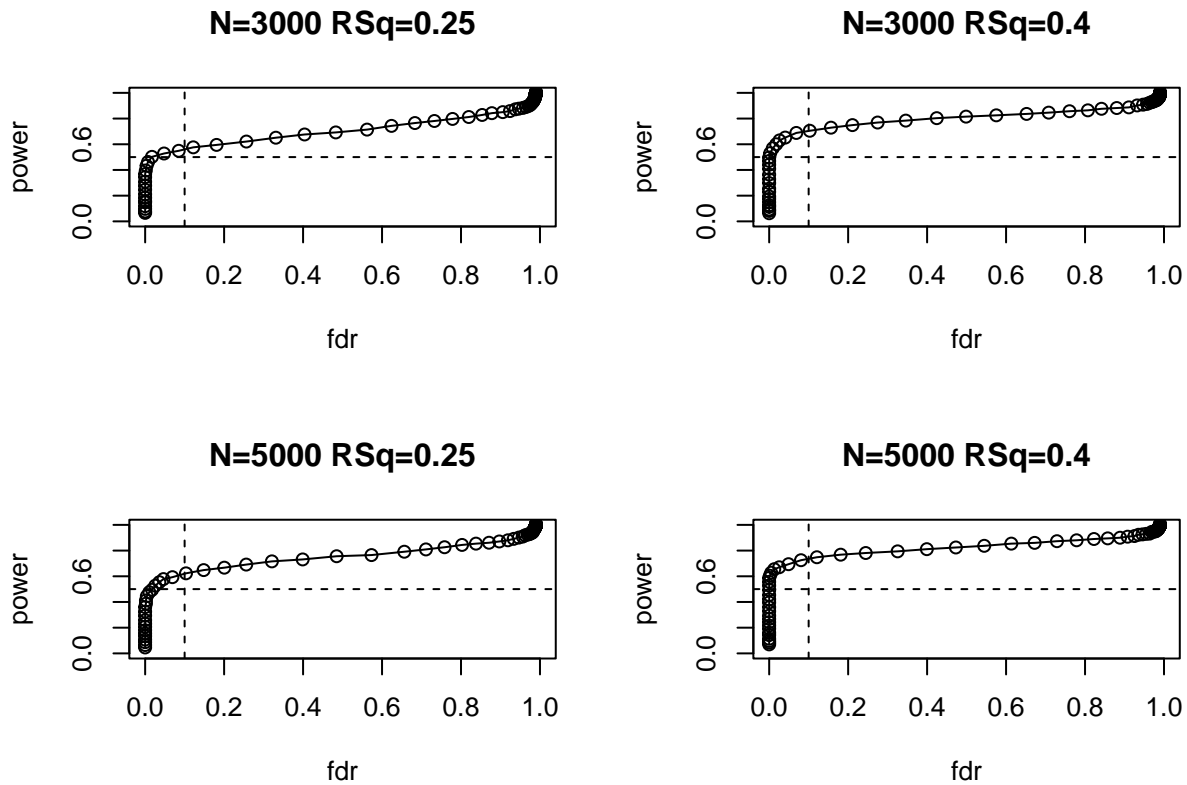
    signal=X[,causal]*%*%rexp(q)
    noise=rnorm(n,sd=sqrt(var(signal)*(1-RSq)/RSq))
    y=signal+noise

    # Fitting lasso
    fm=glmnet(y=y,x=X,lambda=lambda)
    B=as.matrix(fm$beta)

    # Evaluating power and fdr
    for(i in 1:length(lambda)){
      discovery_set=which(fm$beta[,i]!=0)
      FDP[i,mcRep]=mean(!(which(fm$beta[,i]!=0)%in%causal))
      PWR[i,mcRep]=mean(causal%in%discovery_set)
    }
  }
  power=rowMeans(PWR,na.rm=TRUE)
  fdr=rowMeans(FDP,na.rm=TRUE)
  tmp=which.min(abs(power-.5))
  SCEN$FDRPower50[scen]=fdr[tmp]

  tmp=which.min(abs(fdr-.1))
  SCEN$PowerFDR10[scen]=power[tmp]
```

```
plot(power~fdr,type='o',main=paste0('N=',n,' RSq=',RSq),xlim=c(0,1),ylim=c(0,1));abline(v=.1,lty=2);a
}
```



**Question 2:** For each scenario report:

- 2.1) The FDR you need to tolerate to achieve a power of at least 0.5.
- 2.2) The power for an FDR smaller or equal than 0.1.

```
print(round(SCEN,4))
```

```
##      RSq      n FDRPower50 PowerFDR10
## 1 0.25 3000      0.0173      0.549
## 2 0.40 3000      0.0000      0.705
## 3 0.25 5000      0.0175      0.623
## 4 0.40 5000      0.0000      0.725
```