

R package

multiLocalFDR

최 정 인

01



Functions in multiLocalFDR

multiLocalFDR

Overview

multiLocalFDR is a package for multi-dimensional local-FDR estimation using a semiparametric mixture method. The two pillars of the proposed approach are Efron's empirical null principle and log-concave density estimation for the alternative distribution. A unique feature of our method is that it can be extended to compute the local false discovery rates by combining multiple lists of p-values.

- `SpMix()` provides estimates parameters of null and alternative distribution of our semiparametric mixture method.
- `FDR()` provides estimates of FDR or local-FDR for given lists of z-values / p-values.
- `plotFDR()` plots estimated semiparametric mixture distribution and provides threshold z-value for null and alternative distribution.

You can learn more about them in `vignette("multiLocalFDR")`.

Semiparametric Mixture Density Estimation for given z-values

Description

`spMix` returns localFDR estimates and semiparametric mixture density estimates for given multi-dimensional lists of z-values, which are the probit-transformed p-values. For the hypothesis testing `spMix` uses a two-component semiparametric mixture model to estimate the localFDR from the z-values. The two pillars of the proposed approach are Efron's empirical null principle and log-concave density estimation for the alternative distribution.

Arguments

| | |
|------------------------|---|
| <code>z</code> | Matrix which column indicates z-values, probit-transformed p-values. |
| <code>tol</code> | Stopping criteria for the EM algorithm. If maximum absolute difference of current and previous gamma value is smaller than <code>tol</code> , i.e. $\max_i \gamma_i^{(k+1)} - \gamma_i^{(k)} < tol$, for k-th step, then optimization stops. (default: 5e-6) |
| <code>leftNull</code> | If TRUE, a null distribution is placed to the left of the alternative distribution. (default: TRUE) |
| <code>max_iter</code> | Maximum number of iterations in the EM algorithm. (default: 30) |
| <code>mono</code> | If TRUE, localFDR is in ascending order of z-values. (default: TRUE) |
| <code>thre_z</code> | Threshold value which only z-values smaller than <code>thre.z</code> are used to compute the log-concave estimates <code>f_1</code> in M-step. |
| <code>Uthre_gam</code> | Upper threshold of gamma which are used to compute stopping criteria for the EM algorithm. |
| <code>Lthre_gam</code> | Lower threshold of gamma which are used to compute stopping criteria for the EM algorithm. |

Value

Estimates of semiparametric mixture model for `f` for given z-values.

| | |
|-----------------------|--|
| <code>p0</code> | Prior probability for null distribution |
| <code>mu0 sig0</code> | Parameter estimates of normal null distribution, $N(\mu_0, \sigma_0^2)$ |
| <code>f</code> | Probability estimates of semiparametric mixture model for each z-value point. |
| <code>f1</code> | Probability estimates of alternative distribution of mixture model for each z-value point. |
| <code>localFDR</code> | localFDR estimates for given z-values |
| <code>iter</code> | Number of iterations of EM algorithm to compute localFDR. |

```
if (leftNull) {
  q0 <- quantile(z, probs = .9)
  p0 <- mean(z <= q0)
  mu0 <- mean(z[z <= q0])
  sig0 <- sd(z[z <= q0])
  f0 <- dmvnorm(z, mu0, sig0)
  mu1 <- mean(z[z > q0])
  sig1 <- sd(z[z > q0])
  f1 <- dnorm(z, mu1, sig1)
}
else {
  q0 <- quantile(z, probs = .7)
  p0 <- mean(z >= q0)
  mu0 <- mean(z[z >= q0])
  sig0 <- sd(z[z >= q0])
  f0 <- dmvnorm(z, mu0, sig0)
  mu1 <- mean(z[z < q0])
  sig1 <- sd(z[z < q0])
  f1 <- dnorm(z, mu1, sig1)
}
```

Different initialization

FDR estimation for given z-values / p-values

Description

`FDR` returns FDR estimates for given multi-dimensional lists of z-values. `FDR` also provides `localFDR` estimates for both z-values and p-values. `FDR` imports `SpMix` for a two-component semiparametric mixture model to estimate the FDR / `localFDR` from the z-values / p-values.

Arguments

| | |
|------------------------|--|
| <code>z</code> | Matrix which column indicates z-values, probit-transformed p-values. |
| <code>tol</code> | Stopping criteria for the EM algorithm. If maximum absolute difference of current and previous gamma value is smaller than <code>tol</code> , i.e. $\max_j \gamma_j^{(k+1)} - \gamma_j^{(k)} < \text{tol}$, for k-th step, then optimization stops. (default: 5e-6) |
| <code>p_value</code> | If TRUE, input are p-values. If FALSE, input are z-values. (default: FALSE) |
| <code>local</code> | If TRUE, <code>FDR</code> returns <code>localFDR</code> estimates for given z-values or p-values. If FALSE, <code>FDR</code> returns FDR estimates. (default: FALSE) |
| <code>leftNull</code> | If TRUE, a null distribution is placed to the left of the alternative distribution. (default: TRUE) |
| <code>max_iter</code> | Maximum number of iterations in the EM algorithm. (default: 30) |
| <code>mono</code> | If TRUE, FDR is in ascending order of z-values. (default: TRUE) |
| <code>thre_z</code> | Threshold value which only z-values smaller than <code>thre.z</code> are used to compute the log-concave estimates <code>f_1</code> in M-step. |
| <code>Uthre_gam</code> | Upper threshold of gamma which are used to compute stopping criteria for the EM algorithm. |
| <code>Lthre_gam</code> | Lower threshold of gamma which are used to compute stopping criteria for the EM algorithm. |

Value

Estimates of FDR or `localFDR` for given z-values / p-values.

| | |
|-----------------------|---|
| <code>FDR</code> | FDR estimates for given z-values / p-values |
| <code>localFDR</code> | local FDR estimates for given z-values / p-values |

p-value or z-value? Default: z-values

localFDR or FDR? Default: FDR

leftNull or RightNull? Default: leftNull

```
FDR <- function(z, tol = 5e-6, p_value = FALSE, local = FALSE, leftNull = TRUE, max_iter
{
  if (p_value) {
    if (leftNull) {
      z = qnorm(1-z)
    }
    else {
      z = qnorm(z)
    }
  }

  SpMixParams <- SpMix(z, tol, max_iter, leftNull, mono, thre_z, Uthre_gam, Lthre_gam)

  if (local) {
    return(SpMixParams$localFDR)
  }
  else {
    # todo
    # localFDR => FDR
    return(SpMixParams$localFDR)
  }
}
```

plotFDR

Description

`plotFDR` returns `plotFDR` estimates and semiparametric mixture density estimates for given multi-dimensional lists of z-values, which are the probit-transformed p-values. For the hypothesis testing `plotFDR` uses a two-component semiparametric mixture model to estimate the `plotFDR` from the p-values. The two pillars of the proposed approach are Efron's empirical null principle and log-concave density estimation for the alternative distribution.

Usage

```
plotFDR(z, p0, mu0, sig0, f1, localFDR, leftNull = TRUE, thre_localFDR = 0.2)
```

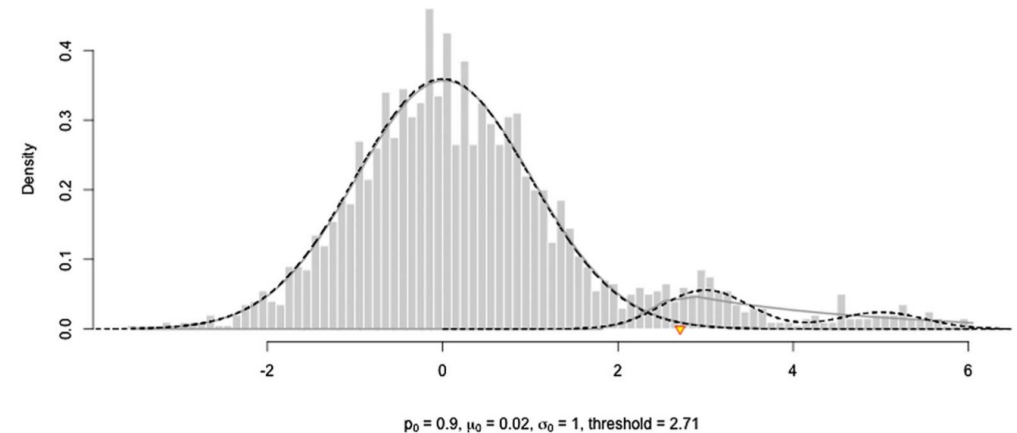
Arguments

| | |
|----------------------------|---|
| <code>z</code> | Matrix which column indicates z-values, probit-transformed p-values. |
| <code>p0</code> | Prior probability for null distribution |
| <code>mu0</code> | <code>sig0</code> Parameter estimates of normal null distribution, $N(\mu_0, \text{sig0}^2)$ |
| <code>f1</code> | Probability estimates of alternative distribution of mixture model for each z-value point. |
| <code>localFDR</code> | <code>localFDR</code> estimates for given z-values |
| <code>leftNull</code> | If TRUE, a null distribution is placed to the left of the alternative distribution. (default: TRUE) |
| <code>thre_localFDR</code> | Threshold of <code>localFDR</code> for null and alternative distribution (default: 0.2) |

Value

Plot estimated semiparametric mixture density and return threshold value.

`thre` Threshold z-value for null and alternative distribution



02

Data for Case-studies

 main ▾ [multiLocalFDR / data /](#)

JungiinChoi add rda data

..



Carina.rda

add rda data



Microarrays.rda

add rda data



Pathways.rda

add rda data

```
R> library(multiLocalFDR)
```

```
# z-values (probit transformed p-values) and p-values from Carina dataset
```

```
R> z <- Carina$z
```

```
R> p <- Carina$p
```

```
# get the parameter estimates of null and alternative distribution
```

```
R> SpMix_Carina <- SpMix(z, leftNull = FALSE)
```

```
# get FDR and local-FDR estimates using z-values
```

```
R> FDR_z <- FDR(z, leftNull = FALSE)
```

```
R> localFDR_z <- FDR(z, local = TRUE, leftNull = FALSE)
```

```
# get FDR and local-FDR estimates using p-values
```

```
R> FDR_p <- FDR(p, p_value = TRUE, leftNull = FALSE)
```

```
R> localFDR_p <- FDR(p, p_value = TRUE, local = TRUE, leftNull = FALSE)
```

```
# plot density estimates and threshold for null and alternative distribution
```

```
R> plotFDR(z, SpMix_Carina$p0, SpMix_Carina$mu0, SpMix_Carina$sig0,  
           SpMix_Carina$f1, SpMix_Carina$localFDR, leftNull = FALSE)
```


03

To-do

multiLocalFDR

Overview

multiLocalFDR is a package for multi-dimensional local-FDR estimation using a semiparametric mixture method. The two pillars of the proposed approach are Efron's empirical null principle and log-concave density estimation for the alternative distribution. A unique feature of our method is that it can be extended to compute the local false discovery rates by combining multiple lists of p-values.

- `SpMix()` provides estimates parameters of null and alternative distribution of our semiparametric mixture method.
- `FDR()` provides estimates of FDR or local-FDR for given lists of z-values / p-values.
- `plotFDR()` plots estimated semiparametric mixture distribution and provides threshold z-value for null and alternative distribution.

You can learn more about them in `vignette("multiLocalFDR")`.

1. `localFDR()` `FDR()` 함수 분리
2. `FDR()` 구현
3. `SpMix()` 1-d인 경우 3-component (two-sided) option 추가
4. Galaxy radial velocity data update

8

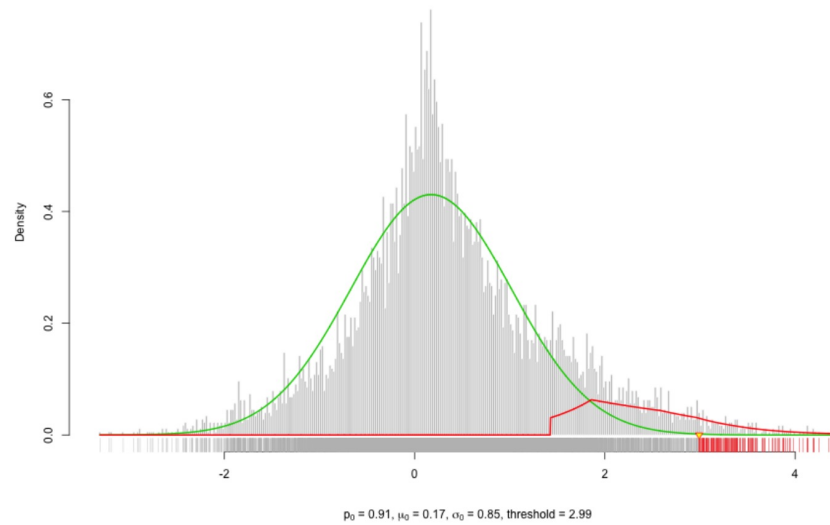
multiLocalFDR

Figure 2:

- 4. 논문 Simulation Plot update
- 5. 논문 내용 작성