MHT Simulation

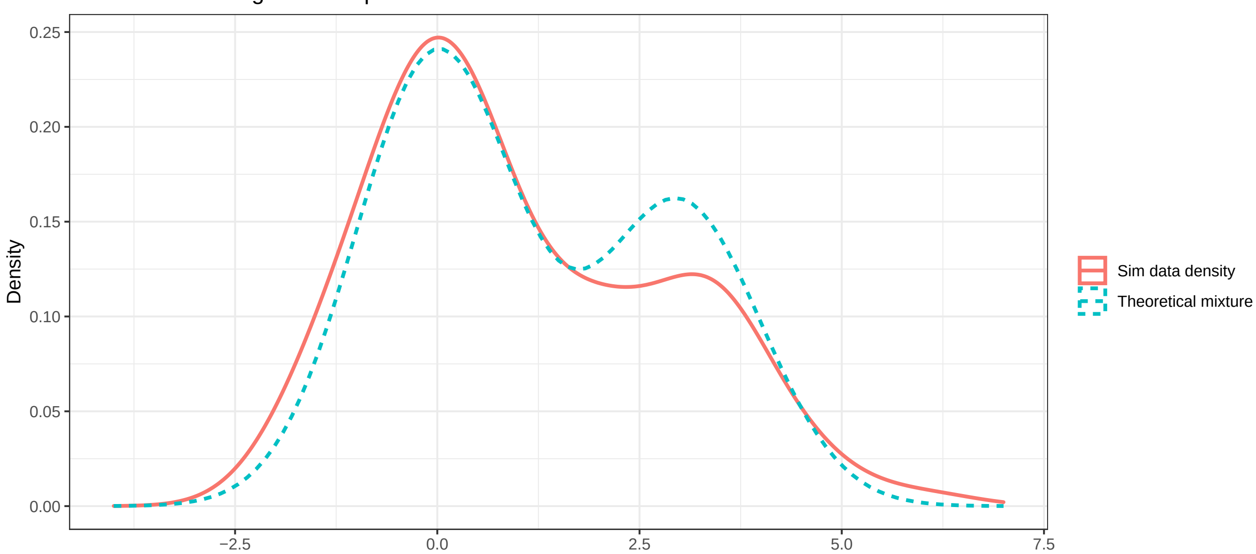
Connor Demorest

9/4/2021

Background

To generate fake data that have a known *pi*, the proportion of true alternative hypotheses (or false null hypotheses), m random $Bernoulli(\pi)$ variables are created. This gives the "truth" statement for each alternative hypothesis. The null hypotheses are distributed Normal(0, 1) and the alternative hypotheses have $Normal(\mu, 1)$ distribution, creating a bi-modal mixture distribution with the density in each mode depending on the value of *pi*, shown in Figure 1.

Figure 1: Empirical vs theoretical mixture of Normal distributions



m=100 simulations, 60% from N(0,1) and 40% from N(3,1) distributions.

After creating the mixture distribution, the $Pr(X > x|\mu = 0)$ (the p-value) is calculated. The p-values are corrected for multiplicity errors using the Bonferroni method and the Benjamini-Hochberg (BH) method. The unadjusted (no correction) and the Bonferroni and BH corrected p-values are used to determine the proportion of rejections of the null hypothesis when the null hypothesis is false (power) and the proportion of incorrect rejections of the null hypothesis (the Type 1 error) divided by the the total proportion of rejections of the null hypothesis (called "discoveries" from now on). This proportion calculated by $FDP = \frac{\text{\# incorrect rejections}}{\text{Total \# rejections}}$ is the false discovery proportion (FPD). The expected value of the FDP in the long run is the false discovery rate (FDR).

Simulation studies for BH and Lfdr

Study 1: Benjamini-Hochberg method compared to Bonferroni correction and no adjustment on FDR, independent case

1) Generate θ_i for $i = 1, 2, ..., m$ from an iid random sample of $Ber(\pi_1)$, where π_1 is the probability of $\theta = 1$, and π_0 is the probability of $\theta = 0$, $\pi_0 + \pi_1 = 1$.
2) For each $\theta_i = 1$, draw from a $f_1 = N(\mu, 1)$ and for $\theta_i = 0$ draw from $f_0 = N(0, 1)$. Let $X_i = \theta_i * f_1 + (1 - \theta_i) * f_0$.
3) Calculate the p-value $Pr(X > x|\mu = 0)$ for each X_i testing $H_0 : \mu = 0$ against $H_0 : \mu > 0$, and apply each correction procedure at $\alpha = 0.05$ level. Track the proportion of rejections of the null hypothesis when the null hypothesis is true (false discovery proportion) and the proportion of rejecting the null hypothesis when the null hypothesis is false (Power).
4) Repeat 1-3 50 times for each of the correction methods, for values of $\pi_1 = 0.01\ to\ 0.99$, $m = \{30, 100\}$ and $\mu = \{3, 5\}$.

Figure 2: Power of variations of parameters for FDR control methods

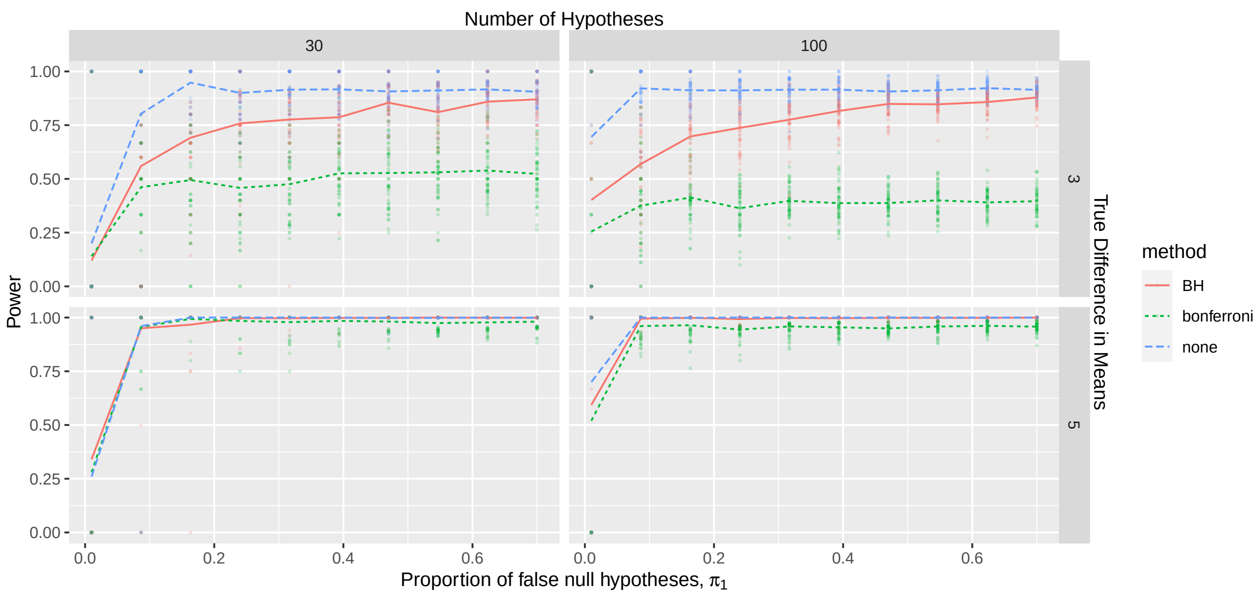
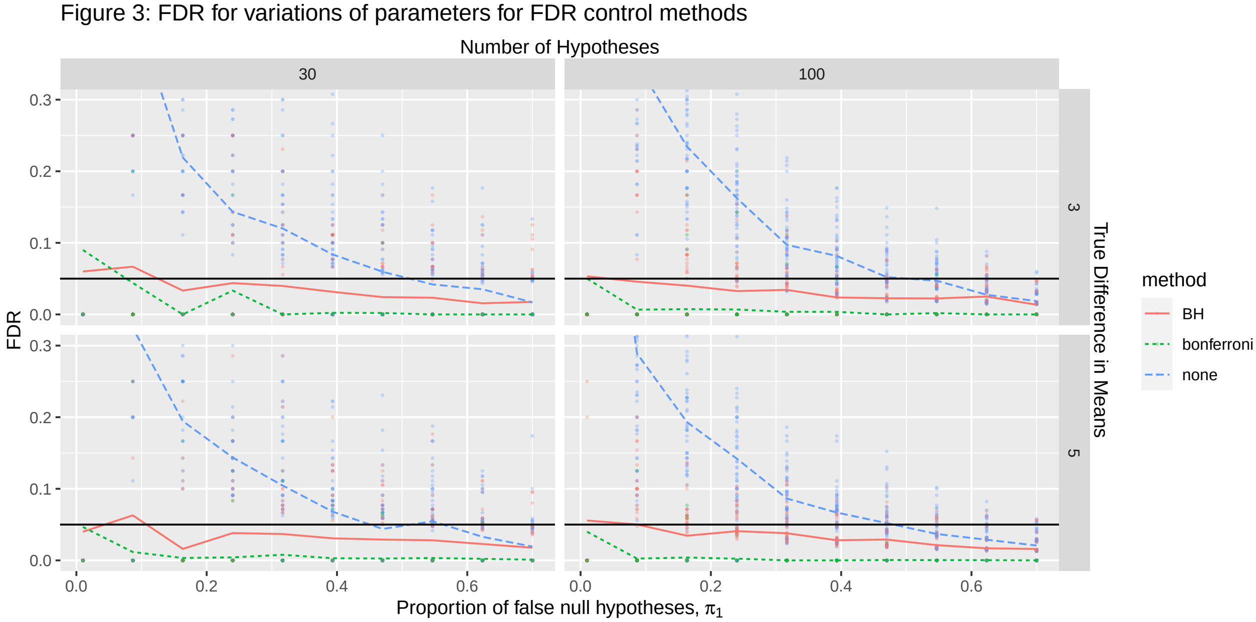


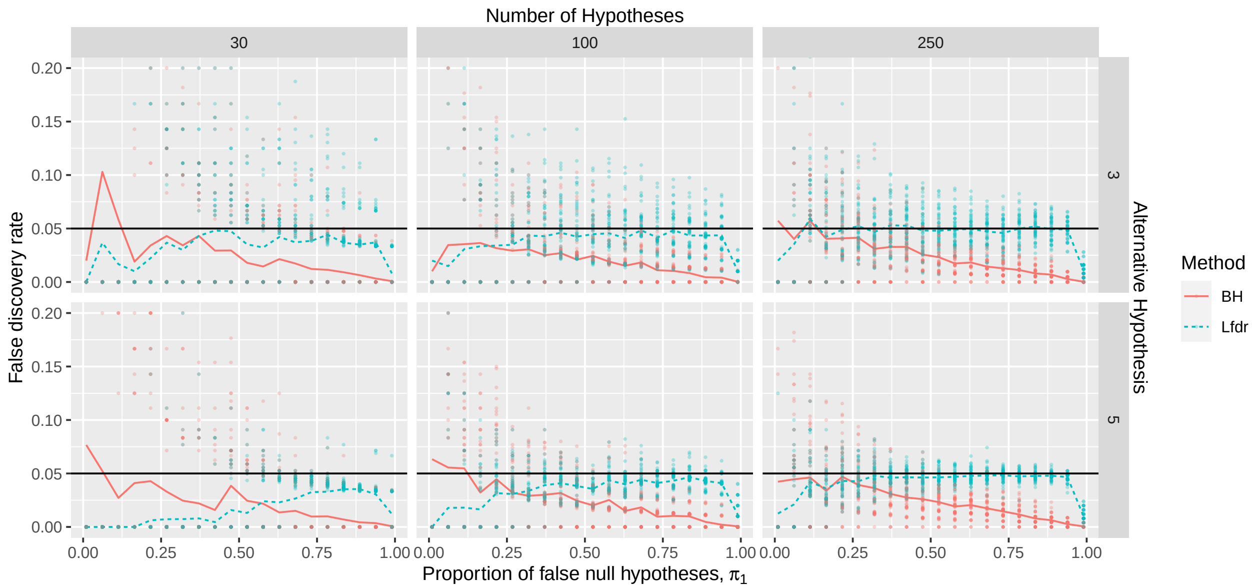
Figure 3: FDR for variations of parameters for FDR control methods



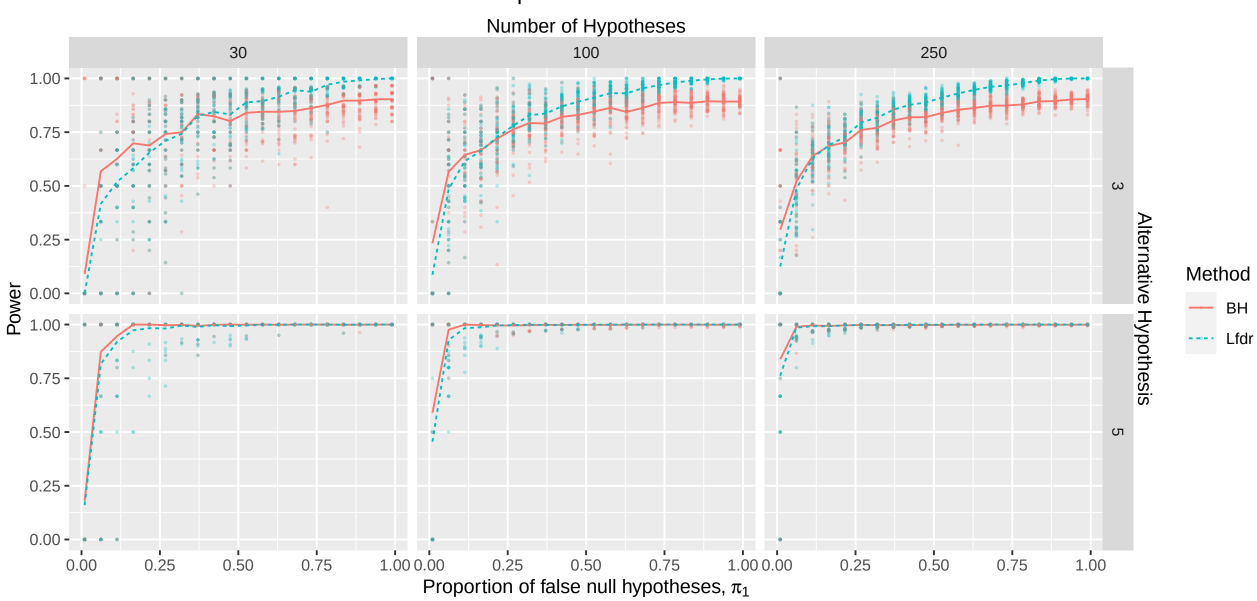
Study 2: Local fdr comparison to BH under PRDS assumption

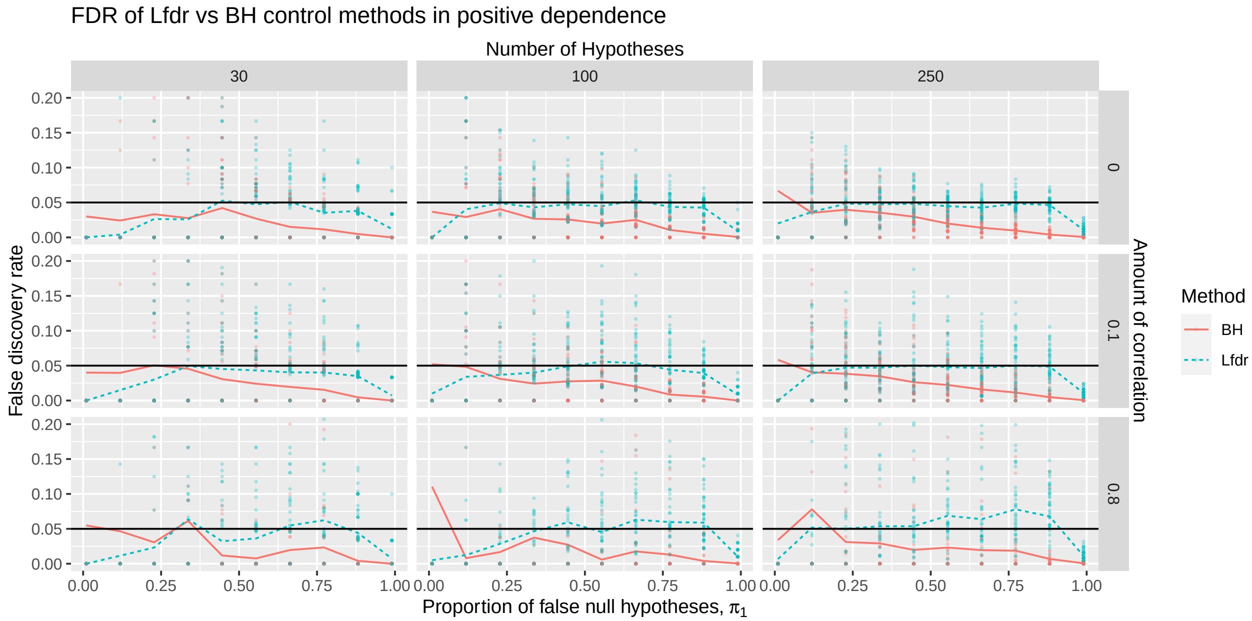
1) Generate θ_i for $i = 1, 2, ..., m$ from an iid random sample of $Ber(\pi_1)$, where π_1 is the probability of $\theta = 1$, and π_0 is the probability of $\theta = 0$, $\pi_0 + \pi_1 = 1$.
2) Sample a vector X of length m draws from a multivariate Normal distribution with mean μ when $\theta_i = 1$ and mean 0 when $\theta_i = 0$, and covariance matrix of 1 on the diagonals and ρ on the off-diagonals. Let $f(X) = \pi_0 * f_0(X) + \pi_1 * f_1(X)$ be the "mixture density", where $f_0(X) = N(0,1)$ and $f_1(X) = N(\mu, 1)$.
3) Calculate the p-value $Pr(X > x|\mu = 0)$ for each X_i and adjust using the BH method.
4) Calculate the $lfdr(X) = \pi_0 * f_0(X)/f(X)$ for each X_i.
5) Order the $lfdr(X)$ from smallest to largest, reject the null hypothesis for the first k ordered hypotheses where $\frac{1}{k}\sum_{i=1}^{k} lfdr(X_{(i)}) < \alpha$ and all BH adjusted p-values $p < \alpha$. Track the proportion of rejections of the null hypothesis when the null hypothesis is true (false discovery proportion) and the proportion of rejecting the null hypothesis when the null hypothesis is false (Power) for both Local FDR and BH FDR adjustment methods.
6) Repeat 1-5 for 50 replicates for values of $\pi_1 = 0.01$ *to* 0.99, $rho = 0$ (independent case), $\mu = \{3, 5\}$, $m = \{30, 100, 250\}$, and again for values of $\pi_1 = 0.01$ *to* 0.7, $rho = \{0, 0.1, 0.8\}$ (positive dependent case), $\mu = 3$, $m = \{30, 100, 250\}$.

FDR of Lfdr vs BH control methods in independence

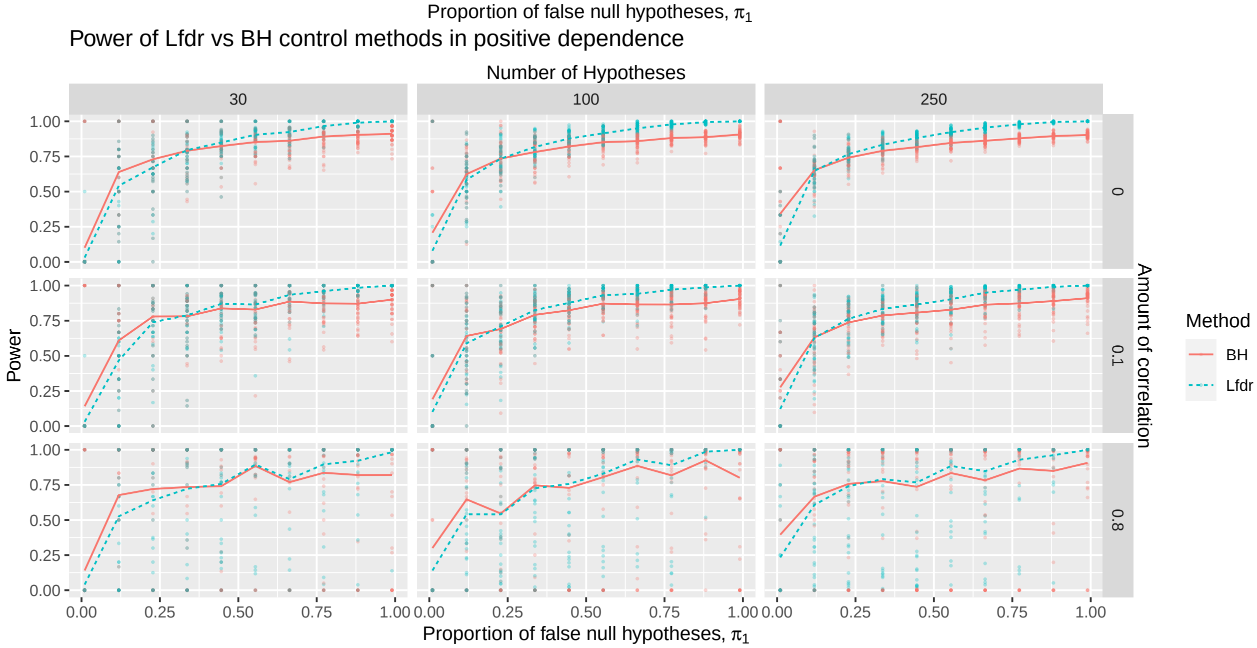


Power of Lfdr vs BH control methods in independence

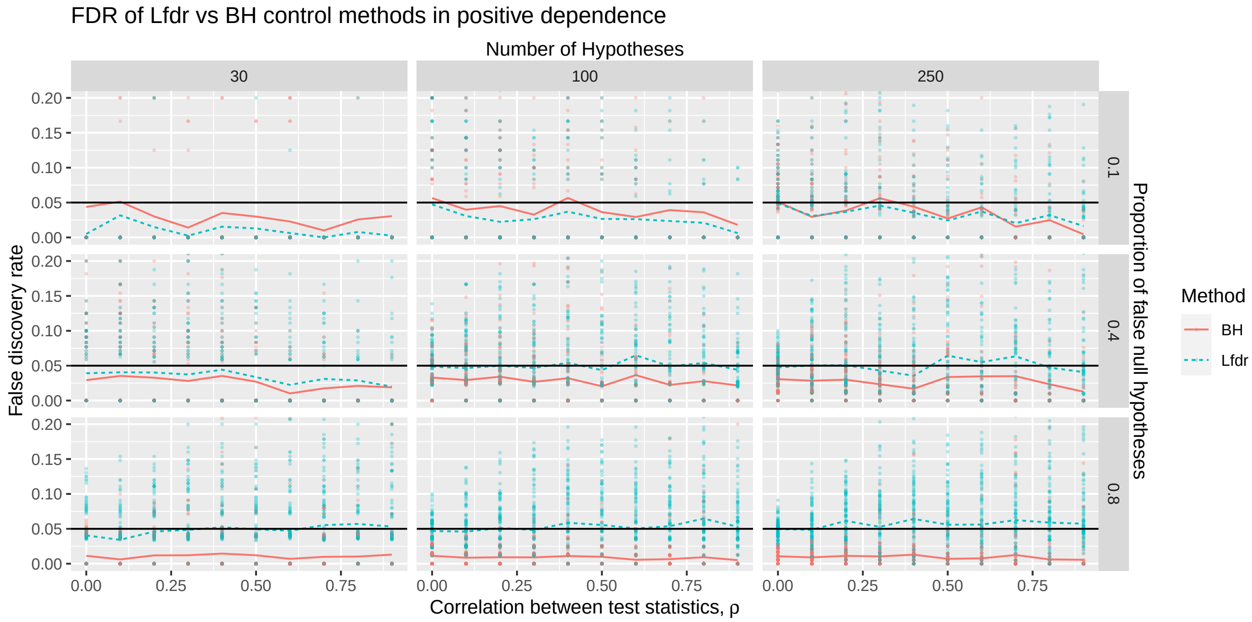


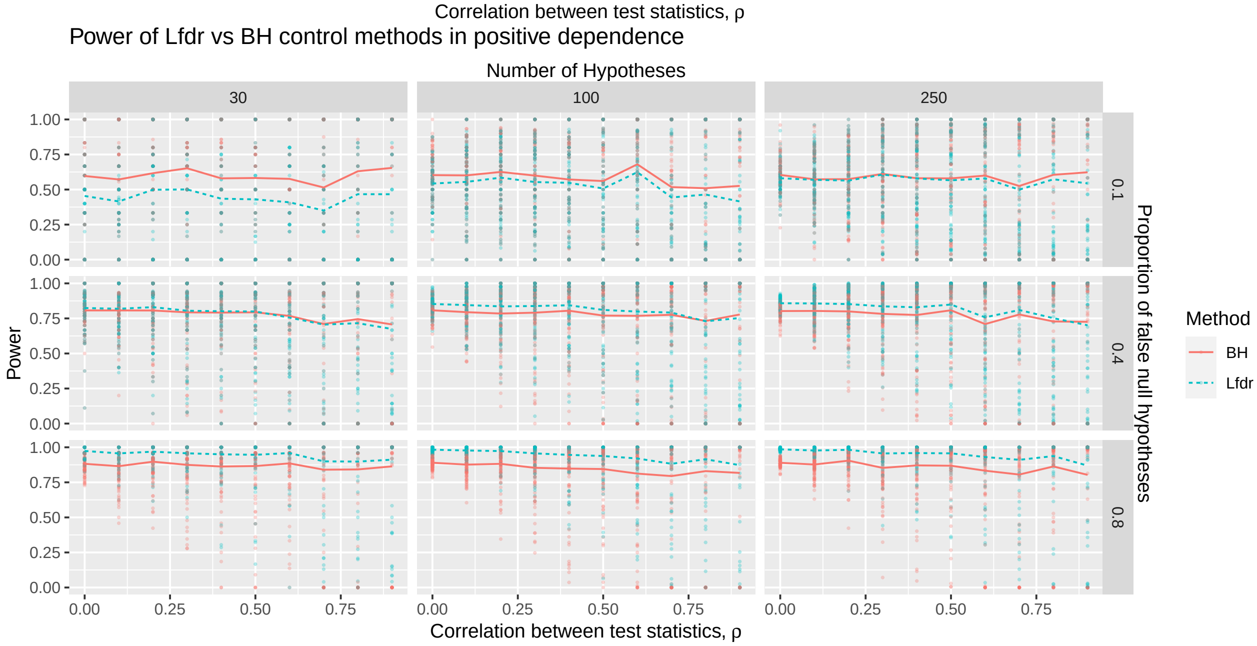


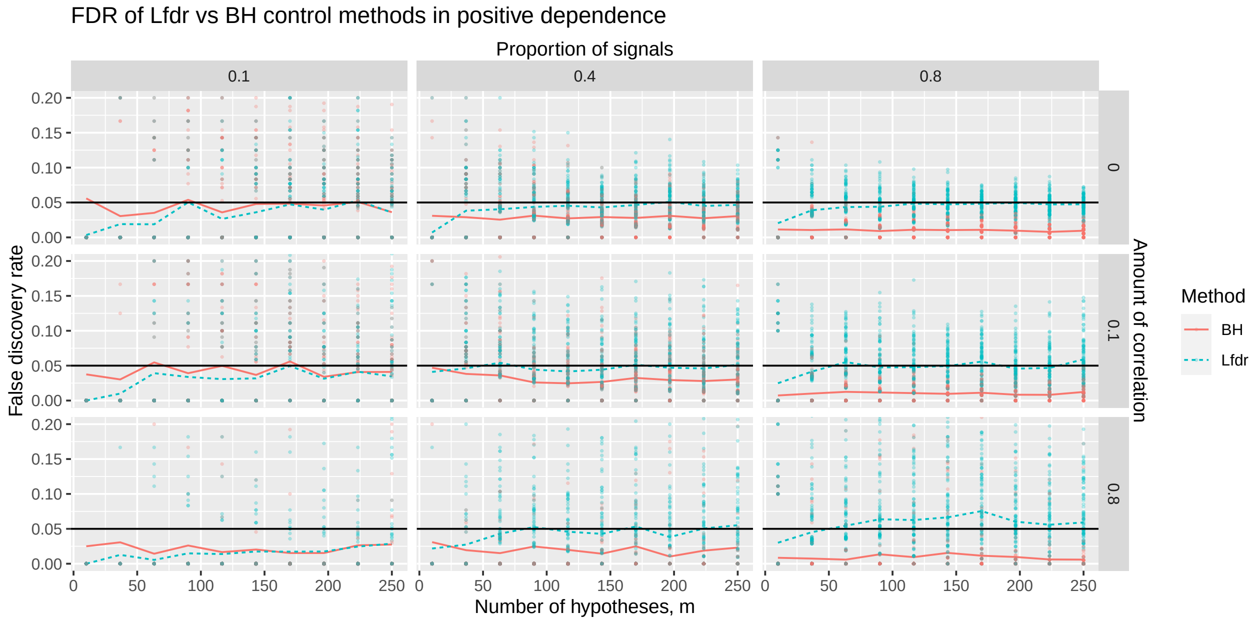
FDR of Lfdr vs BH control methods in positive dependence
Number of Hypotheses
30
100
250
False discovery rate
Amount of correlation
0
0.1
0.8
Proportion of false null hypotheses, π_1
Method
BH
Lfdr

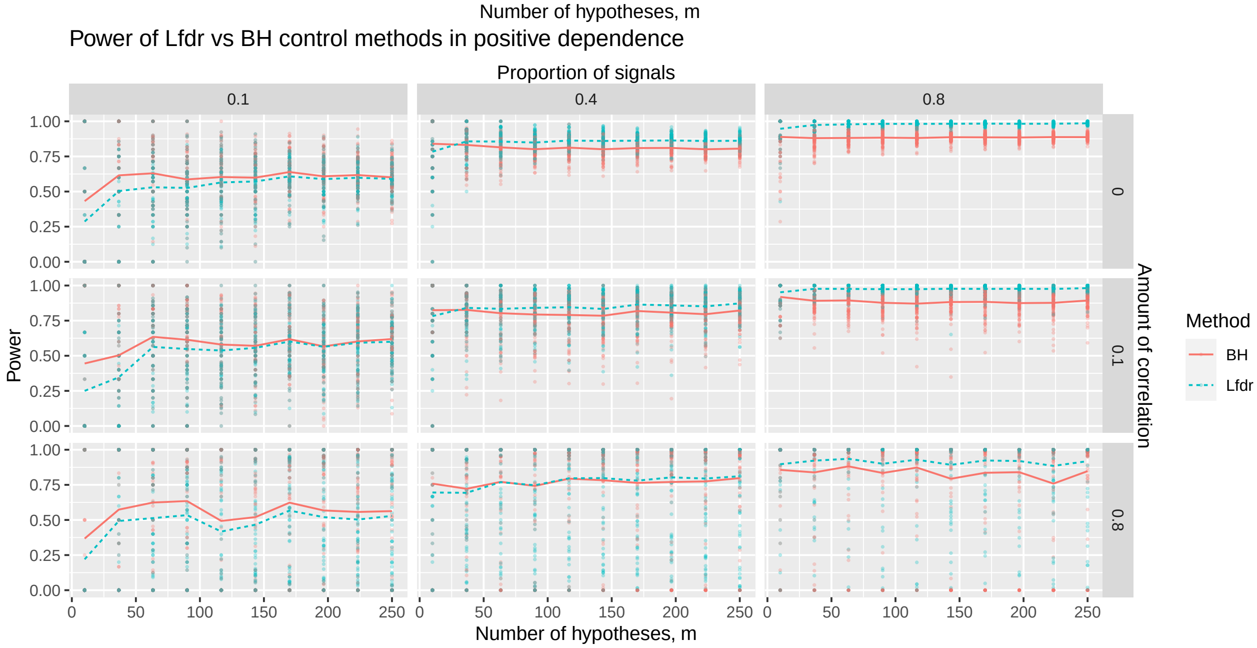


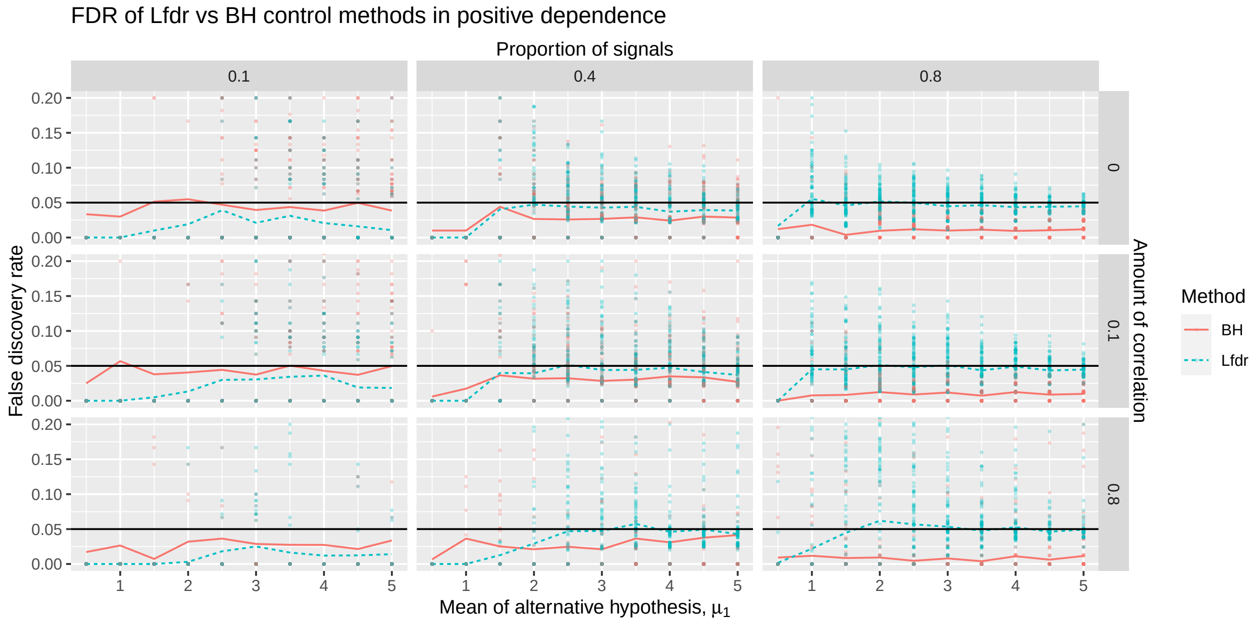
Power of Lfdr vs BH control methods in positive dependence
Number of Hypotheses
30
100
250
Power
Amount of correlation
0
0.1
0.8
Proportion of false null hypotheses, π_1
Method
BH
Lfdr

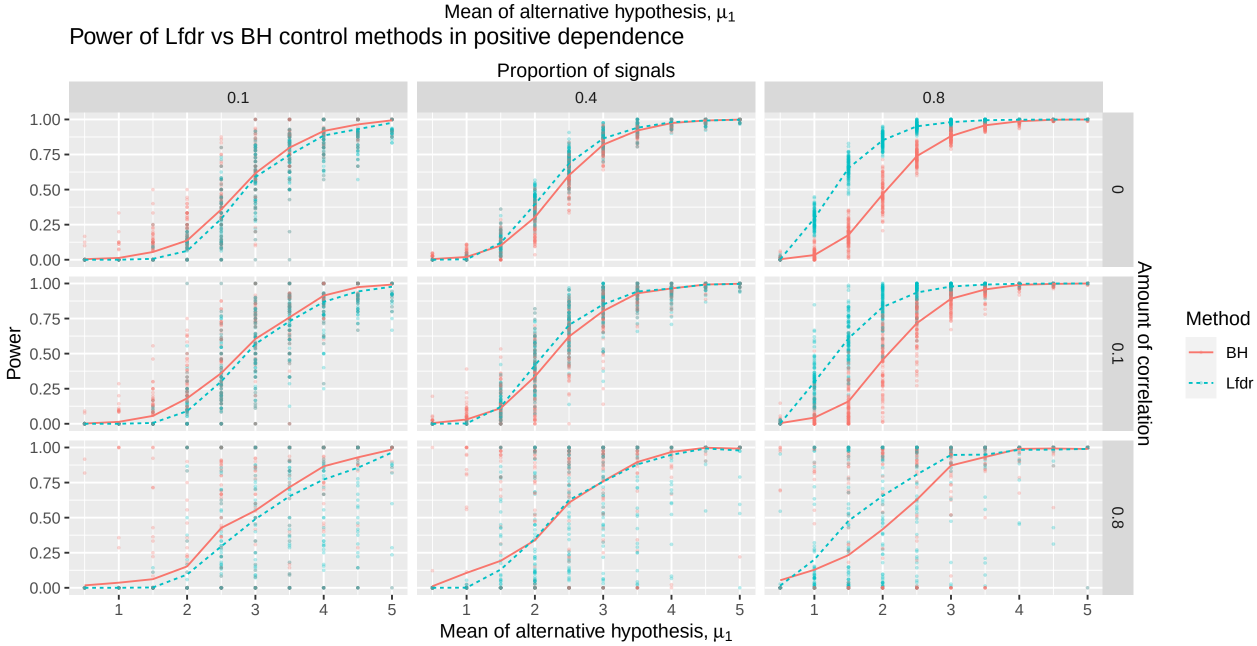

FDR of Lfdr vs BH control methods in positive dependence
Number of Hypotheses
30
100
250
False discovery rate
0.00
0.05
0.10
0.15
0.20
Proportion of false null hypotheses
0.1
0.4
0.8
Correlation between test statistics, ρ
0.00
0.25
0.50
0.75
Method
BH
Lfdr


Power of Lfdr vs BH control methods in positive dependence
Number of Hypotheses
30
100
250
Power
0.00
0.25
0.50
0.75
1.00
Proportion of false null hypotheses
0.1
0.4
0.8
Correlation between test statistics, ρ
Method
BH
Lfdr


FDR of Lfdr vs BH control methods in positive dependence
Proportion of signals
0.1
0.4
0.8
Amount of correlation
0
0.1
0.8
False discovery rate
Number of hypotheses, m
Method
BH
Lfdr


Power of Lfdr vs BH control methods in positive dependence
Proportion of signals
0.1
0.4
0.8
Amount of correlation
0
0.1
0.8
Power
Number of hypotheses, m
Method
BH
Lfdr


FDR of Lfdr vs BH control methods in positive dependence
Proportion of signals
0.1
0.4
0.8
False discovery rate
0.20
0.15
0.10
0.05
0.00
1
2
3
4
5
Mean of alternative hypothesis, μ_1
Amount of correlation
0
0.1
0.8
Method
BH
Lfdr


Power of Lfdr vs BH control methods in positive dependence
Proportion of signals
0.1
0.4
0.8
Power
1.00
0.75
0.50
0.25
0.00
1
2
3
4
5
Mean of alternative hypothesis, μ_1
Amount of correlation
0
0.1
0.8
Method
BH
Lfdr

Code: