Competitors when Non-null, Varying Sample Size, log2-fold standard deviation, and π_0.

David Gerard

February 10, 2016

Abstract

I compare MOUTHWASH to various competitors. I look at AUC and estimates of π_0. I also look at the Kendall's Tau between the p-values or lfdr's for the different methods. This file runs the same analysis as varyNsampNullpiLog2sd.pdf but with smaller sample sizes and smaller log2sd.

1 Competitors

For each competitor, I performed two methods. The first method consisted of a two-step procedure:

1. Estimate $\hat{\beta}_{[2,i]}$ and it's corresponding standard error $\hat{s}_i$.
2. Run ASH on $\hat{\beta}_{[2,i]}$ and $\hat{s}_i$.

The second method was to just calculate a normal (or t, where appropriate) p-values from $\hat{\beta}_{[2,i]}/\hat{s}_i$.

The ASH methods provide an estimate of π_0. I obtained an estimate of π_0 from the p-values by the `qvalue` package in `R` [Storey, 2002]. In some cases for the quasi-binomial methods, the largest p-values were less than 0.9 and `qvalue` would return an error (because it uses the largest p-values to estimate the proportion of nulls). For these, I used the upper quartile of p-values to estimate the proportion of nulls. Maybe a bad idea.

The methods that Mengyin and I have coded to get $\hat{\beta}_{[2,i]}$ and $\hat{s}_i$ were

- VOOM [Law et al., 2014].
- RUVseq [Risso et al., 2014] followed by VOOM [Law et al., 2014] with the estimated confounding factors. Half of the factors were used as control genes. RUVseq is RUV2 on the log(counts + 1) matrix.
- SVAseq [Leek, 2014] followed by VOOM [Law et al., 2014] with the estimated confounding factors. SVAseq is SVA on the log(counts + 1) matrix.
- Quasi-binomial glm.
- RUVseq + quasi-binomial glm.
- SVAseq + quasi-binomial glm.
- MYRNA, which is just a quasi-binomial glm using the 75th percentile of the samples' counts as covariates [Langmead et al., 2010].
- MYRNA offset, which is just a quasi-binomial glm using the 75th percentile of the samples' counts as offsets [Langmead et al., 2010].
- DESeq2glm [Love et al., 2014].
- The robust regression version of CATE [Wang et al., 2015] on the log(counts + 1).

- The negative controls version of CATE [Wang et al., 2015] on the log(counts + 1).
- SVA [Leek and Storey, 2007] with the number of confounders estimated using the method of Buja and Eyuboglu [1992] on the log(counts + 1), followed by OLS.
- RUV2 [Gagnon-Bartsch et al., 2013] with 50% of the observations being control genes with the number of confounders estimated using the method of Buja and Eyuboglu [1992] on the log(counts + 1), followed by OLS.
- OLS on the log(counts + 1).
- The ridge-regression version of LEAPP [Sun et al., 2012] on the log(counts + 1).
- The soft-thresholding version of LEAPP [Sun et al., 2012] on the log(counts + 1).

Notes:

- "VOOM" means using VOOM [Law et al., 2014] to find weights for each observations, then fitting a linear model using LIMMA [Smyth, 2005].
- LEAPP does not easily provide standard errors, so I excluded it from the ASH analysis. But I still use it for the `qvalue` analysis.
- EdgeR was giving me trouble, so I excluded it.

The factor analysis part of MOUTHWASH was done with the quasi-mle approach of Bai et al. [2012] with the number of hidden confounders using the methods of Buja and Eyuboglu [1992] implemented in the `num.sv()` function in the `sva` package in `R`.

In summary, there are 31 methods that I compared in estimating π_0 and in their AUC:

1. MOUTHWASH
2. voom + ASH
3. Quasi-binomial GLM + ASH
4. MyrnaQB + ASH
5. Myrnaoffqb + ASH
6. RUVseq + voom + ASH
7. SVAseq + voom + ASH
8. RUVseq + Quasi-binomial GLM + ASH
9. SVAseq + Quasi-binomial GLM + ASH
10. DESeq2glm + ASH
11. OLS on log(counts + 1) + ASH
12. RUV2 on log(counts + 1) + ASH
13. SVA on log(counts + 1) + ASH
14. Robust Regression Cate + ASH
15. Negative Control CATE + ASH
16. voom + `qvalue`
17. Quasi-binomial GLM + `qvalue`
18. Myrnaqb + `qvalue`
19. Myrnaoffqb + `qvalue`
20. RUVseq + voom + `qvalue`
21. SVAseq + voom + `qvalue`
22. RUVseq + Quasi-binomial GLM + `qvalue`
23. SVAseq + Quasi-binomial GLM + `qvalue`
24. DESeq2glm + `qvalue`
25. OLS on log(counts + 1) + `qvalue`
26. RUV2 on log(counts + 1) + `qvalue`

27. SVA on log(counts + 1) + `qvalue`
28. Robust Regression CATE + `qvalue`
29. Negative Control CATE + `qvalue`
30. Soft-thresholding version of LEAPP+ `qvalue`
31. Ridge version of LEAPP+ `qvalue`

2 Simulation Study

I ran through 100 repetitions of generating data from GTEX lung data under the following parameter conditions:

- $n \in \{3, 5\}$,
- $p = 1000$,
- $\pi_0 \in \{0.5, 0.9\}$,
- $\sigma_{log2} \in \{1, 5\}$.

I extracted the most expressed p genes (excluding the top 5 expressed genes) from the GTEX lung data and n samples are chosen at random. Half of these samples are randomly given the "treatment" label 1, the other half given the "treatment" label 0. Of the p genes, $\pi_0 p$ were chosen to be non-null. Signal was added by the Poisson-thinning approach in Mengyin's code with a mean log2-fold change of 0 and a standard deviation log2-fold change of σ_{log2}. That is

$$
\begin{aligned}
A_1, \ldots, A_{p/2} &\sim N(0, \sigma_{log2}^2) && (1)\\
B_i &= 2^{A_i} \text{ for } i = 1, \ldots, p/2. && (2)
\end{aligned}
$$

If $A_i > 0$ then we replace $Y_{[1:(n/2),i]}$ with $Binom(Y_{[j,i]}, 1/B_i)$ for $j = 1, \ldots, n/2$. If $A_i < 0$ then we replace $Y_{[(n/2+1):n,i]}$ with $Binom(Y_{[j,i]}, B_i)$ for $j = n/2 + 1, \ldots, n$.

For each iteration, I calculated three things:

1. The pairwise Kendall's tau between the methods' lfdr's or p-values.
2. The AUC using either the lfdrs or p-values.
3. The estimates of π_0.

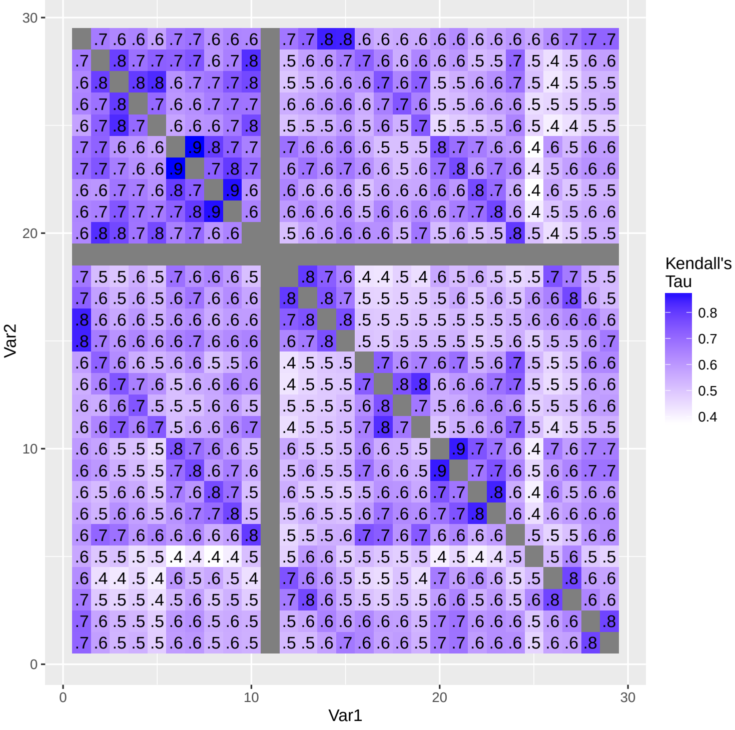
3 Results

For the frequentist procedures, I used the vector of p-values as the predictions and I used the vector of lfdr's from the ASH-like procedures for prediction. These were used to create ROC curves and calculate AUCs. In general, the AUC's were all very similar with the ash-like methods having slightly higher AUC. DESeq2glm (with or without ASH) is the winner when the sample size is 3 vs 3. The results are less clear for 5 vs 5.

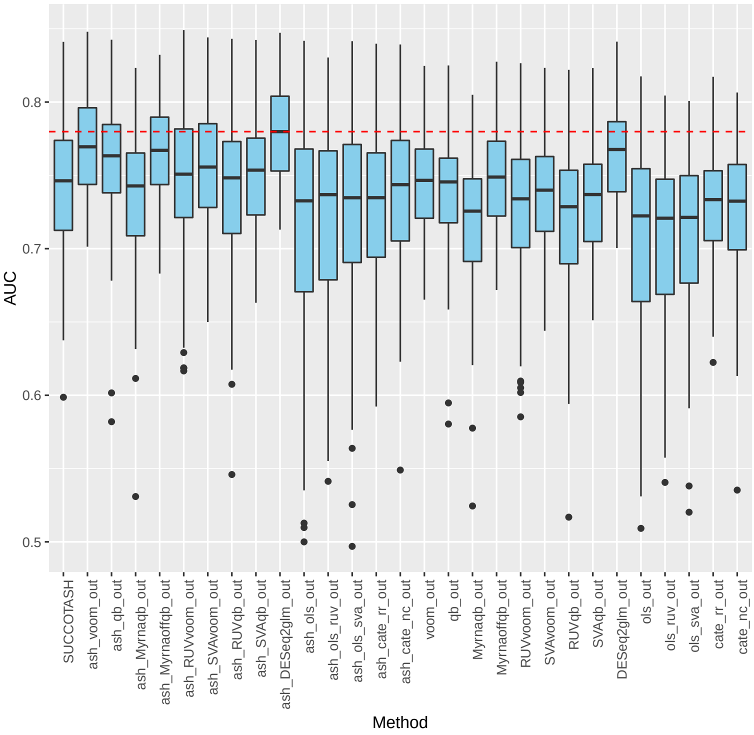
Graphical representations of the median Kendall's taus are presented below. The median Kendall's tau can get quite low between the separate groups — as small as 0.2. This indicates that for many datasets, the rankings can be quite different.

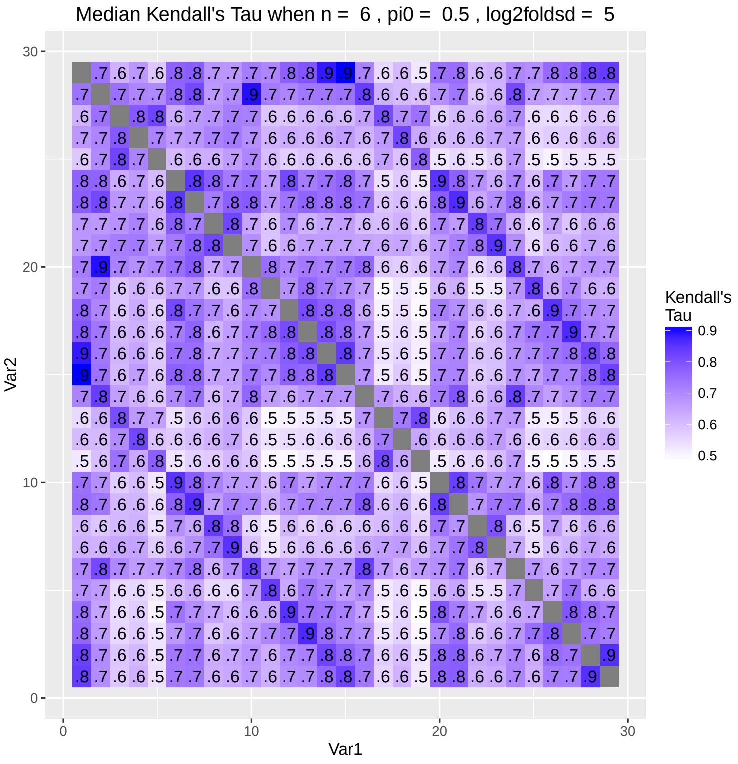
From the p-values, I used the `qvalue` package [Storey, 2002] to estimate π_0. Estimates of π_0 are given from `ashr` for the ASH-like methods. MOUTHWASH (SUCCOTASH) performs the worst in estimating π_0, usually underestimating it. The ASH-like methods usually estimate π_0 to be smaller than their non-ASH counterparts. Notably, the negative controls version of CATE does not excel in the small sample size scenario.

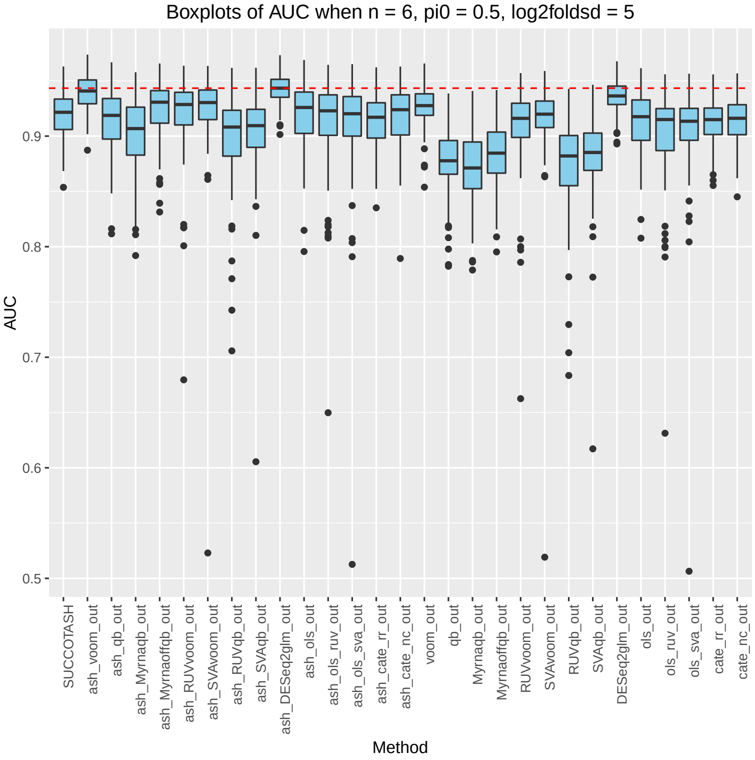
Median Kendall's Tau when n = 6 , pi0 = 0.5 , log2foldsd = 1

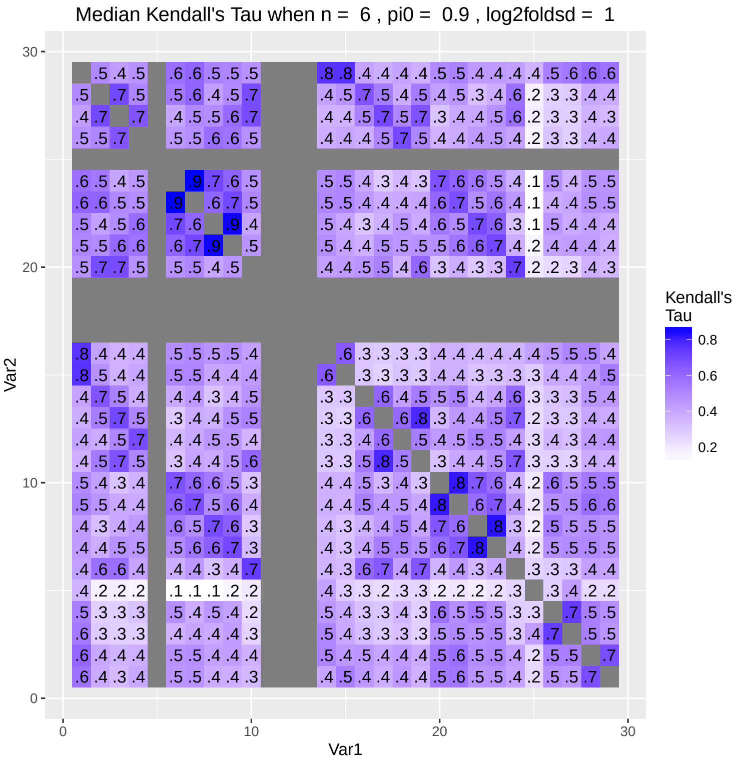


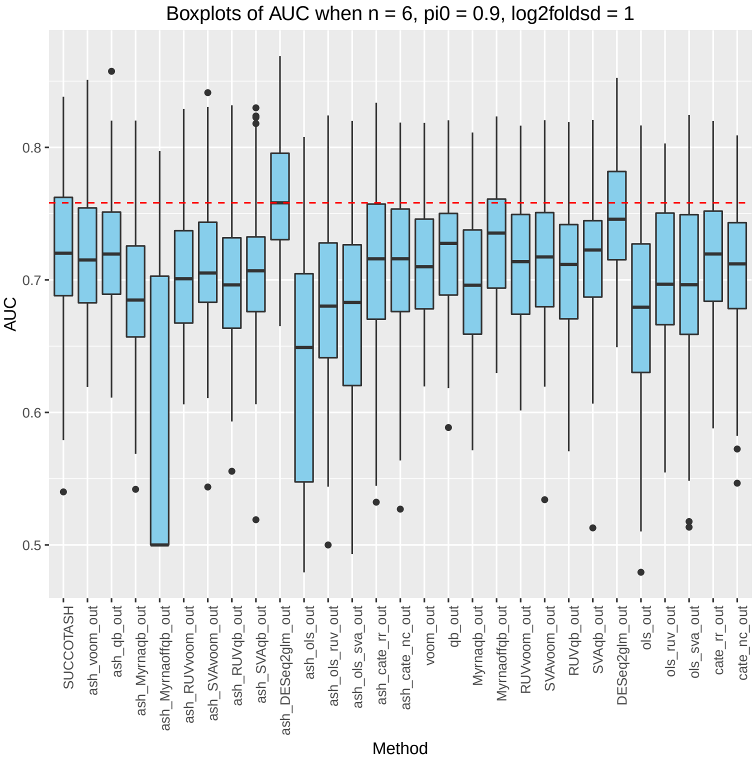
Boxplots of AUC when n = 6, pi0 = 0.5, log2foldsd = 1



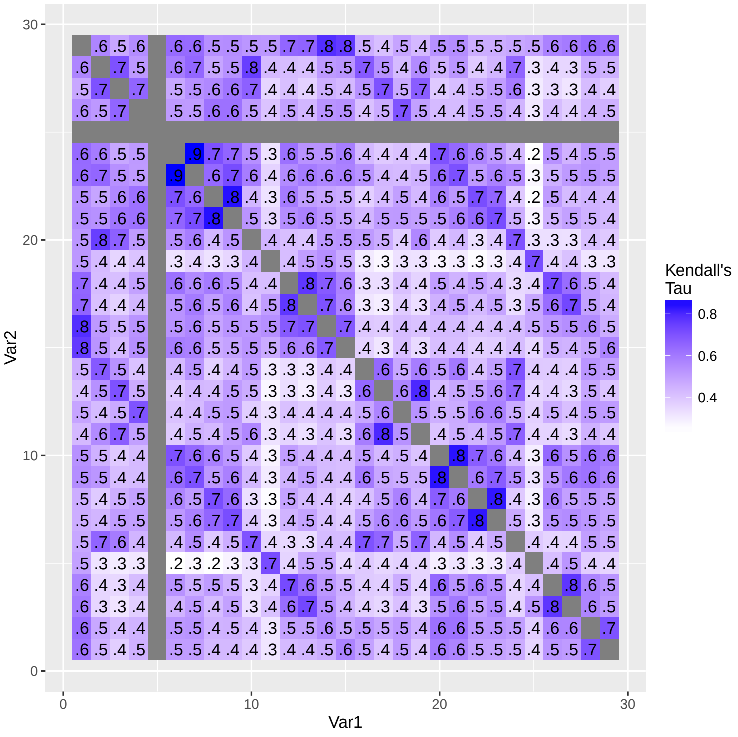

Median Kendall's Tau when n = 6 , pi0 = 0.5 , log2foldsd = 5


Boxplots of AUC when n = 6, pi0 = 0.5, log2foldsd = 5

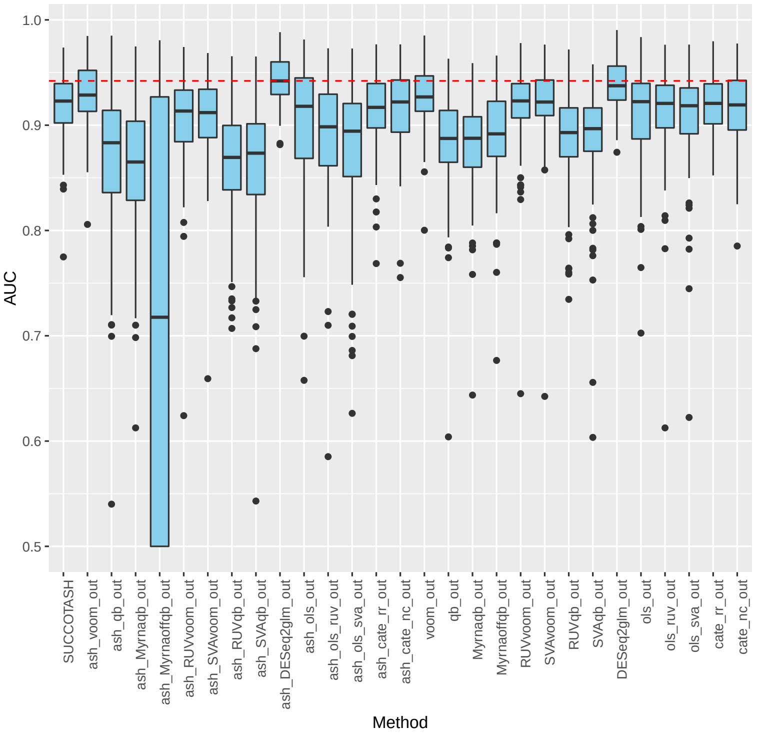

Median Kendall's Tau when n = 6 , pi0 = 0.9 , log2foldsd = 1

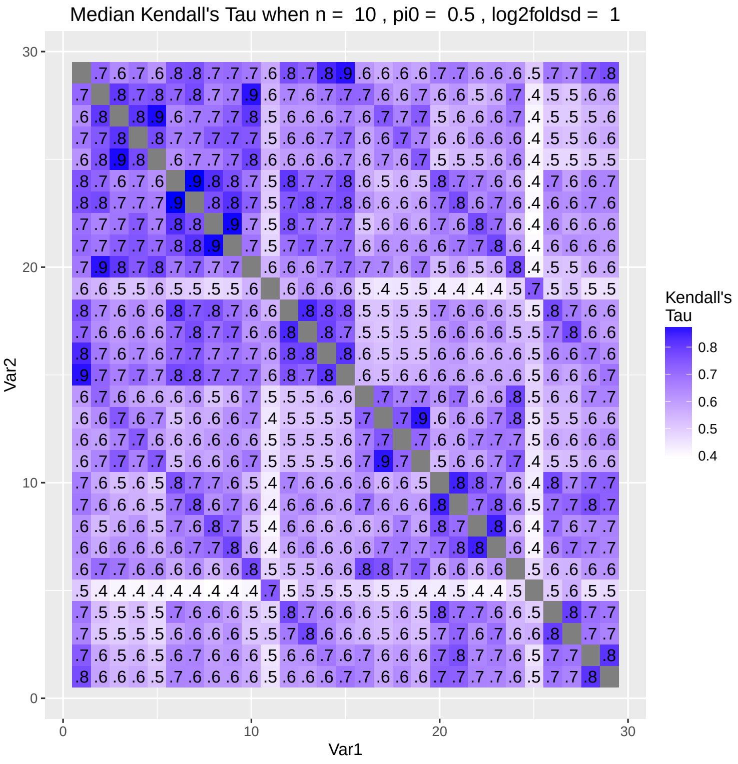

Boxplots of AUC when n = 6, pi0 = 0.9, log2foldsd = 1

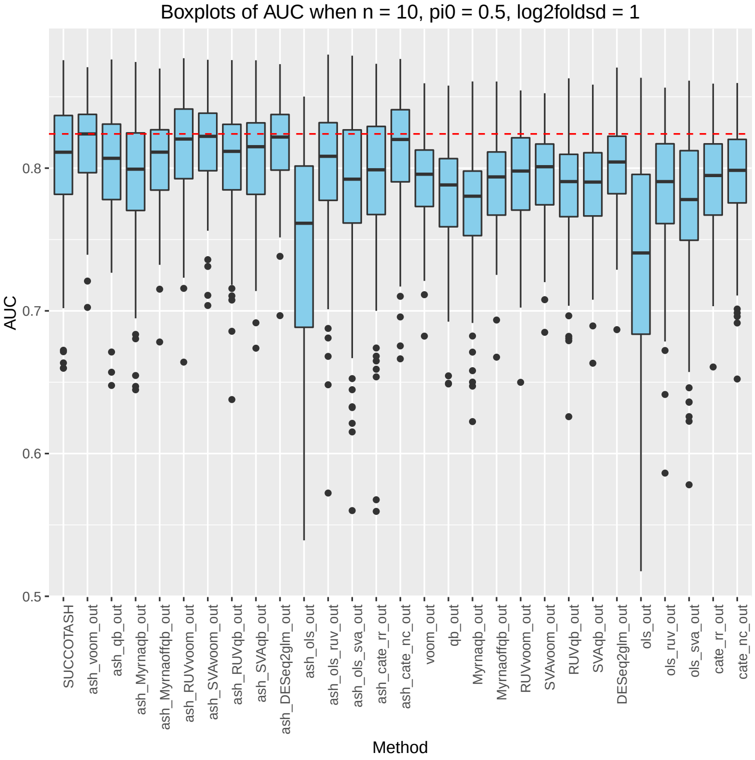
Median Kendall's Tau when n = 6 , pi0 = 0.9 , log2foldsd = 5

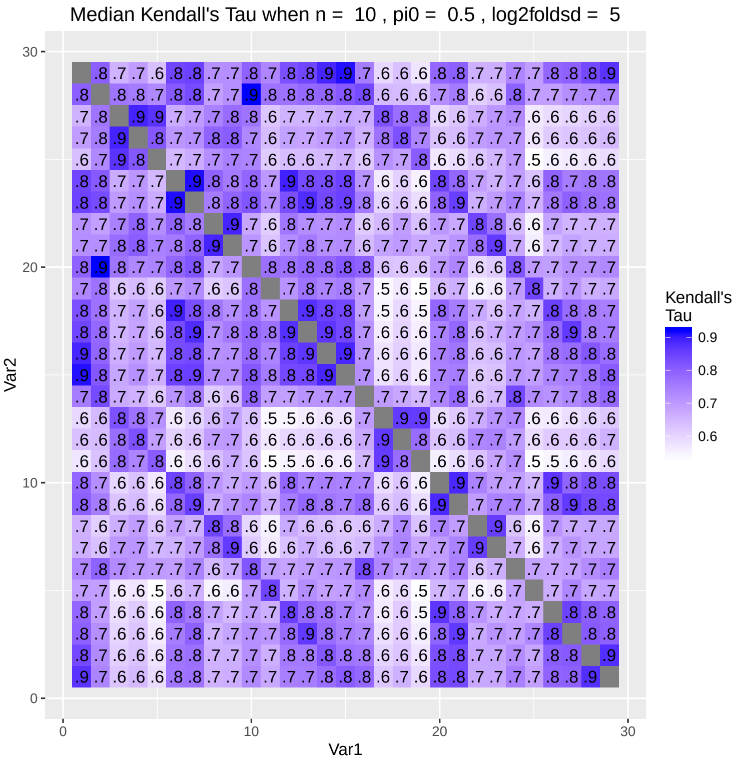


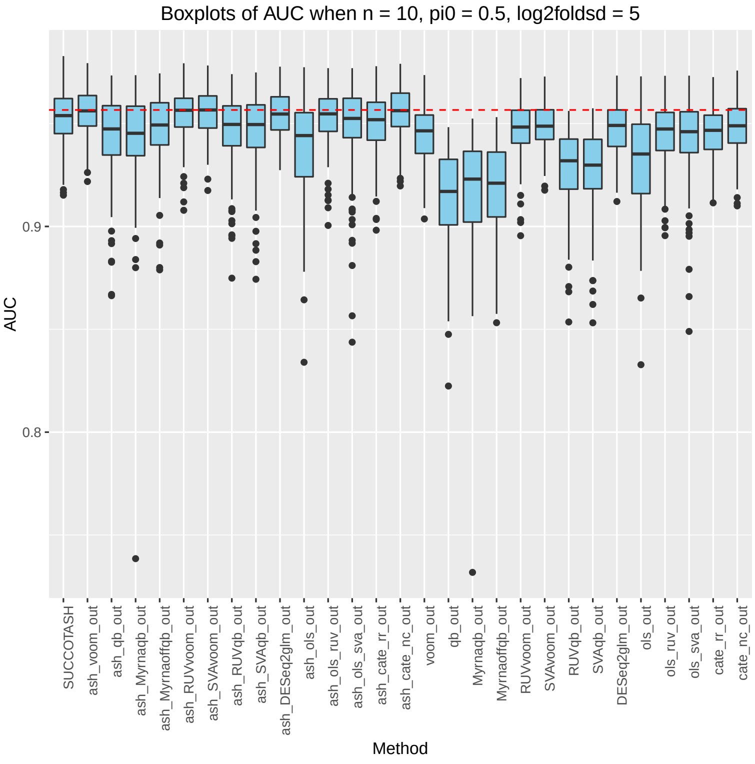
Boxplots of AUC when n = 6, pi0 = 0.9, log2foldsd = 5

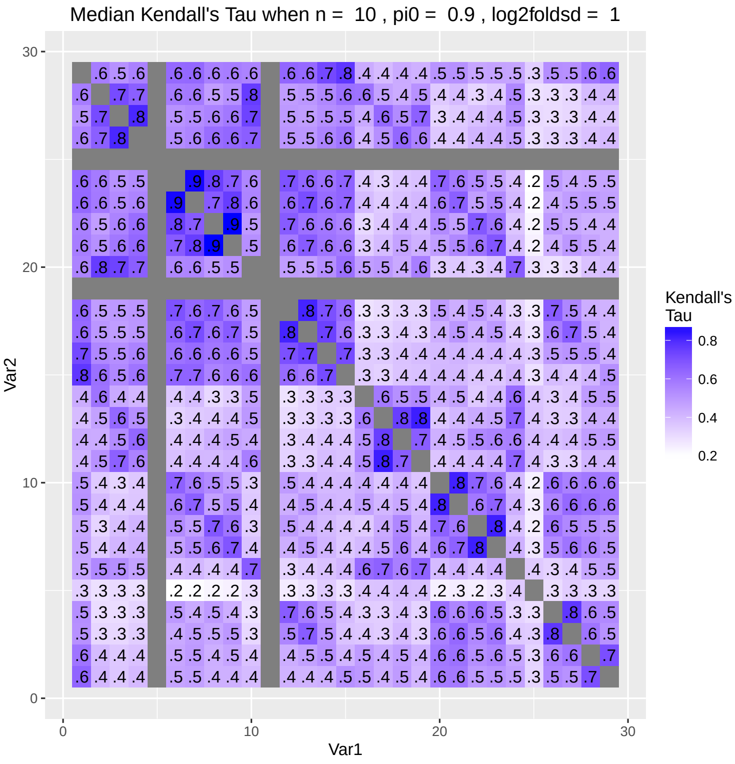



Median Kendall's Tau when n = 10 , pi0 = 0.5 , log2foldsd = 1
.7 .6 .7 .6 .8 .8 .7 .7 .7 .6 .8 .7 .8 .9 .6 .6 .6 .6 .7 .7 .6 .6 .6 .5 .7 .7 .7 .8
.7 .8 .7 .8 .7 .8 .7 .7 .9 .6 .7 .6 .7 .7 .7 .6 .6 .7 .6 .6 .5 .6 .7 .4 .5 .5 .6 .6
.6 .8 .8 .9 .6 .7 .7 .7 .8 .5 .6 .6 .6 .7 .6 .7 .7 .7 .5 .6 .6 .6 .7 .4 .5 .5 .5 .6
.7 .7 .8 .8 .7 .7 .7 .7 .7 .5 .6 .6 .7 .7 .6 .6 .7 .7 .6 .6 .6 .6 .6 .4 .5 .5 .6 .6
.6 .8 .9 .8 .6 .7 .7 .7 .8 .6 .6 .6 .6 .7 .6 .7 .6 .7 .5 .5 .5 .6 .6 .4 .5 .5 .5 .5
.8 .7 .6 .7 .6 .9 .8 .8 .7 .5 .8 .7 .7 .8 .6 .5 .6 .5 .8 .7 .7 .6 .6 .4 .7 .6 .6 .7
.8 .8 .7 .7 .7 .9 .8 .8 .7 .5 .7 .8 .7 .8 .6 .6 .6 .6 .7 .8 .6 .7 .6 .4 .6 .6 .7 .6
.7 .7 .7 .7 .7 .8 .8 .9 .7 .5 .8 .7 .7 .7 .5 .6 .6 .6 .7 .6 .8 .7 .6 .4 .6 .6 .6 .6
.7 .7 .7 .7 .7 .8 .8 .9 .7 .5 .7 .7 .7 .7 .6 .6 .6 .6 .6 .7 .7 .8 .6 .4 .6 .6 .6 .6
.7 .9 .8 .7 .8 .7 .7 .7 .7 .6 .6 .6 .7 .7 .7 .7 .6 .7 .5 .6 .5 .6 .8 .4 .5 .5 .6 .6
.6 .6 .5 .5 .6 .5 .5 .5 .5 .6 .6 .6 .6 .6 .5 .4 .5 .5 .4 .4 .4 .4 .5 .7 .5 .5 .5 .5
.8 .7 .6 .6 .6 .8 .7 .8 .7 .6 .6 .8 .8 .8 .5 .5 .5 .5 .7 .6 .6 .6 .5 .5 .8 .7 .6 .6
.7 .6 .6 .6 .6 .7 .8 .7 .7 .6 .6 .8 .8 .7 .5 .5 .5 .5 .6 .6 .6 .6 .5 .5 .7 .8 .6 .6
.8 .7 .6 .7 .6 .7 .7 .7 .7 .7 .6 .8 .8 .8 .6 .5 .5 .5 .6 .6 .6 .6 .6 .5 .6 .6 .7 .6
.9 .7 .7 .7 .7 .8 .8 .7 .7 .7 .6 .8 .7 .8 .6 .5 .6 .6 .6 .6 .6 .6 .6 .5 .6 .6 .6 .7
.6 .7 .6 .6 .6 .6 .6 .5 .6 .7 .5 .5 .5 .6 .6 .7 .7 .7 .6 .7 .6 .6 .8 .5 .6 .6 .7 .7
.6 .6 .7 .6 .7 .5 .6 .6 .6 .7 .4 .5 .5 .5 .5 .7 .7 .9 .6 .6 .6 .7 .8 .5 .5 .5 .6 .6
.6 .6 .7 .7 .6 .6 .6 .6 .6 .6 .5 .5 .5 .5 .6 .7 .7 .7 .6 .6 .7 .7 .7 .5 .6 .6 .6 .6
.6 .7 .7 .7 .7 .5 .6 .6 .6 .7 .5 .5 .5 .5 .6 .7 .9 .7 .5 .6 .6 .7 .7 .4 .5 .5 .6 .6
.7 .6 .5 .6 .5 .8 .7 .7 .6 .5 .4 .7 .6 .6 .6 .6 .6 .6 .5 .8 .8 .7 .6 .4 .8 .7 .7 .7
.7 .6 .6 .6 .5 .7 .8 .6 .7 .6 .4 .6 .6 .6 .6 .7 .6 .6 .6 .8 .7 .8 .6 .5 .7 .7 .8 .7
.6 .5 .6 .6 .5 .7 .6 .8 .7 .5 .4 .6 .6 .6 .6 .6 .6 .7 .6 .8 .7 .8 .6 .4 .7 .6 .7 .7
.6 .6 .6 .6 .6 .6 .7 .7 .8 .6 .4 .6 .6 .6 .6 .6 .7 .7 .7 .7 .8 .8 .6 .4 .6 .7 .7 .7
.6 .7 .7 .6 .6 .6 .6 .6 .6 .8 .5 .5 .5 .6 .6 .8 .8 .7 .7 .6 .6 .6 .6 .5 .6 .6 .6 .6
.5 .4 .4 .4 .4 .4 .4 .4 .4 .4 .7 .5 .5 .5 .5 .5 .5 .5 .4 .4 .5 .4 .4 .5 .5 .6 .5 .5
.7 .5 .5 .5 .5 .7 .6 .6 .6 .5 .5 .8 .7 .6 .6 .6 .5 .6 .5 .8 .7 .7 .6 .6 .5 .8 .7 .7
.7 .5 .5 .5 .5 .6 .6 .6 .6 .5 .5 .7 .8 .6 .6 .6 .5 .6 .5 .7 .7 .6 .7 .6 .6 .8 .7 .7
.7 .6 .5 .6 .5 .6 .7 .6 .6 .6 .5 .6 .6 .7 .6 .7 .6 .6 .6 .7 .8 .7 .7 .6 .5 .7 .7 .8
.8 .6 .6 .6 .5 .7 .6 .6 .6 .6 .5 .6 .6 .6 .7 .7 .6 .6 .6 .7 .7 .7 .7 .6 .5 .7 .7 .8
Var2
30
20
10
0
Var1
0
10
20
30
Kendall's
Tau
0.8
0.7
0.6
0.5
0.4


Boxplots of AUC when n = 10, pi0 = 0.5, log2foldsd = 1
AUC
0.8
0.7
0.6
0.5
Method
SUCCOTASH
ash_voom_out
ash_qb_out
ash_Myrnaqb_out
ash_Myrnaoffqb_out
ash_RUVvoom_out
ash_SVAvoom_out
ash_RUVqb_out
ash_SVAqb_out
ash_DESeq2glm_out
ash_ols_out
ash_ols_ruv_out
ash_ols_sva_out
ash_cate_rr_out
ash_cate_nc_out
voom_out
qb_out
Myrnaqb_out
Myrnaoffqb_out
RUVvoom_out
SVAvoom_out
RUVqb_out
SVAqb_out
DESeq2glm_out
ols_out
ols_ruv_out
ols_sva_out
cate_rr_out
cate_nc_out

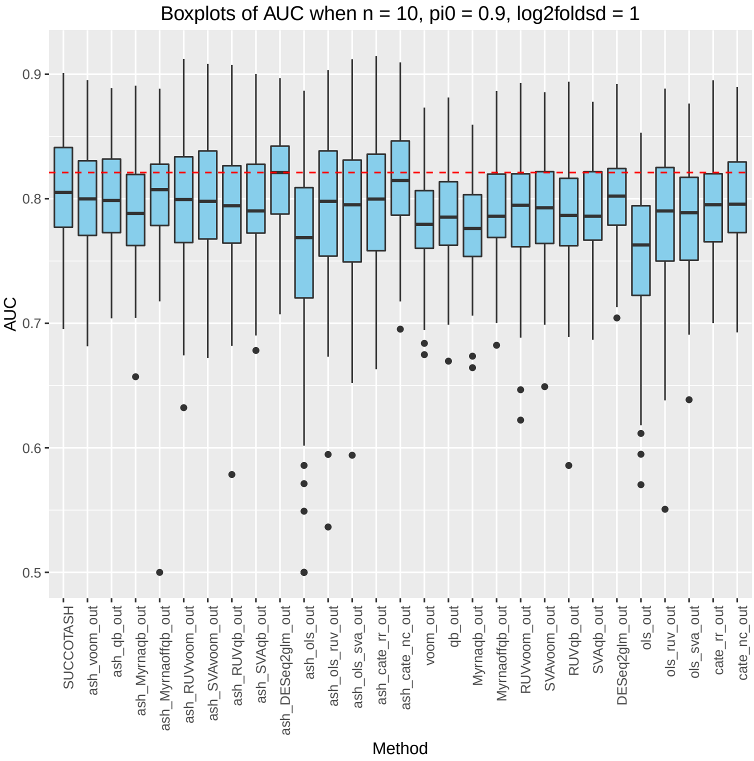

Median Kendall's Tau when n = 10 , pi0 = 0.5 , log2foldsd = 5


Boxplots of AUC when n = 10, pi0 = 0.5, log2foldsd = 5

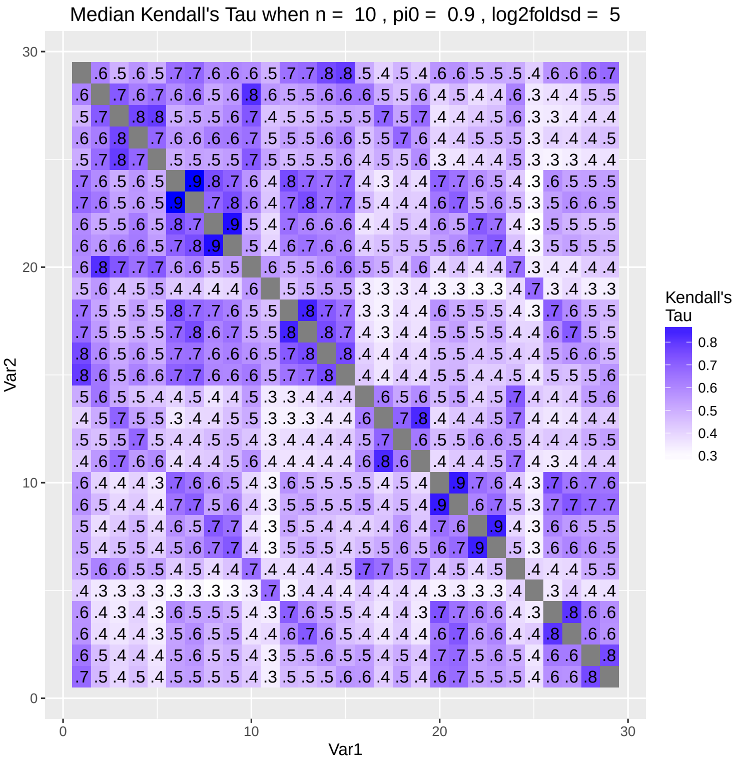


Median Kendall's Tau when n = 10 , pi0 = 0.9 , log2foldsd = 1

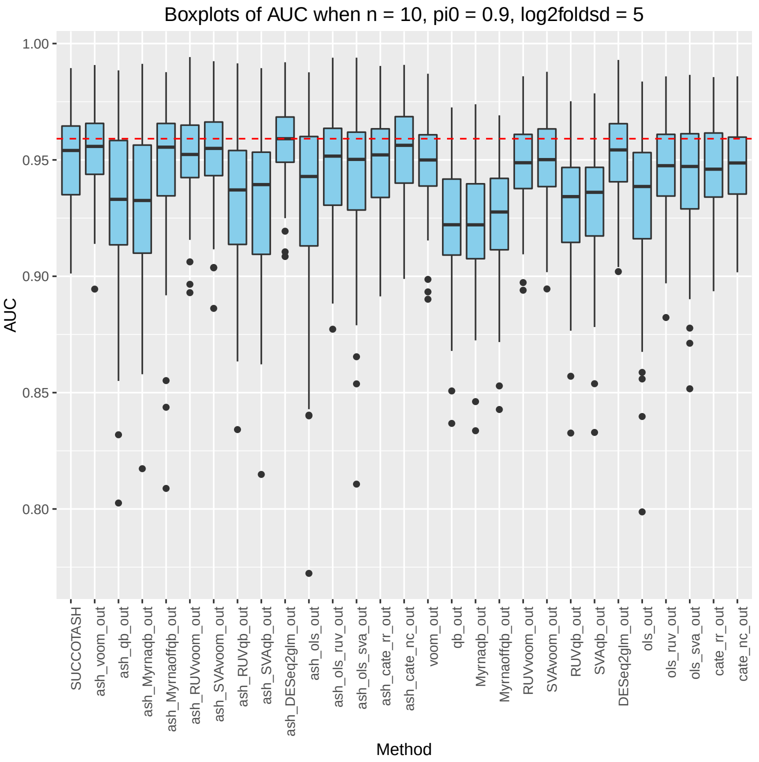
	.6	.5	.6		.6	.6	.6	.6	.6		.6	.6	.7	.8	.4	.4	.4	.4	.5	.5	.5	.5	.5	.3	.5	.5	.6	.6
.6		.7	.7		.6	.6	.5	.5	.8		.5	.5	.5	.6	.6	.5	.4	.5	.4	.4	.3	.4	.5	.3	.3	.3	.4	.4
.5	.7		.8		.5	.5	.6	.6	.7		.5	.5	.5	.5	.4	.6	.5	.7	.3	.4	.4	.4	.5	.3	.3	.3	.4	.4
.6	.7	.8			.5	.6	.6	.6	.7		.5	.5	.6	.6	.4	.5	.6	.6	.4	.4	.4	.4	.5	.3	.3	.3	.4	.4
.6	.6	.5	.5			.9	.8	.7	.6		.7	.6	.6	.7	.4	.3	.4	.4	.7	.6	.5	.5	.4	.2	.5	.4	.5	.5
.6	.6	.5	.6		.9		.7	.8	.6		.6	.7	.6	.7	.4	.4	.4	.4	.6	.7	.5	.5	.4	.2	.4	.5	.5	.5
.6	.5	.6	.6		.8	.7		.9	.5		.7	.6	.6	.6	.3	.4	.4	.4	.5	.5	.7	.6	.4	.2	.5	.5	.4	.4
.6	.5	.6	.6		.7	.8	.9		.5		.6	.7	.6	.6	.3	.4	.5	.4	.5	.5	.6	.7	.4	.2	.4	.5	.5	.4
.6	.8	.7	.7		.6	.6	.5	.5			.5	.5	.5	.6	.5	.5	.4	.6	.3	.4	.3	.4	.7	.3	.3	.3	.4	.4
.6	.5	.5	.5		.7	.6	.7	.6	.5			.8	.7	.6	.3	.3	.3	.3	.5	.4	.5	.4	.3	.3	.7	.5	.4	.4
.6	.5	.5	.5		.6	.7	.6	.7	.5		.8		.7	.6	.3	.3	.4	.3	.4	.5	.4	.5	.4	.3	.6	.7	.5	.4
.7	.5	.5	.6		.6	.6	.6	.6	.5		.7	.7		.7	.3	.3	.4	.4	.4	.4	.4	.4	.4	.3	.5	.5	.5	.4
.8	.6	.5	.6		.7	.7	.6	.6	.6		.6	.6	.7		.3	.3	.4	.4	.4	.4	.4	.4	.4	.3	.4	.4	.4	.5
.4	.6	.4	.4		.4	.4	.3	.3	.5		.3	.3	.3	.3		.6	.5	.5	.4	.5	.4	.4	.6	.4	.3	.4	.5	.5
.4	.5	.6	.5		.3	.4	.4	.4	.5		.3	.3	.3	.3	.6		.8	.8	.4	.4	.4	.5	.7	.4	.3	.3	.4	.4
.4	.4	.5	.6		.4	.4	.4	.5	.4		.3	.4	.4	.4	.5	.8		.7	.4	.5	.5	.6	.6	.4	.4	.4	.5	.5
.4	.5	.7	.6		.4	.4	.4	.4	.6		.3	.3	.4	.4	.5	.8	.7		.4	.4	.4	.4	.7	.4	.3	.3	.4	.4
.5	.4	.3	.4		.7	.6	.5	.5	.3		.5	.4	.4	.4	.4	.4	.4	.4		.8	.7	.6	.4	.2	.6	.6	.6	.6
.5	.4	.4	.4		.6	.7	.5	.5	.4		.4	.5	.4	.4	.5	.4	.5	.4	.8		.6	.7	.4	.3	.6	.6	.6	.6
.5	.3	.4	.4		.5	.5	.7	.6	.3		.5	.4	.4	.4	.4	.4	.5	.4	.7	.6		.8	.4	.2	.6	.5	.5	.5
.5	.4	.4	.4		.5	.5	.6	.7	.4		.4	.5	.4	.4	.4	.5	.6	.4	.6	.7	.8		.4	.3	.5	.6	.6	.5
.5	.5	.5	.5		.4	.4	.4	.4	.7		.3	.4	.4	.4	.6	.7	.6	.7	.4	.4	.4	.4		.4	.3	.4	.5	.5
.3	.3	.3	.3		.2	.2	.2	.2	.3		.3	.3	.3	.3	.4	.4	.4	.4	.2	.3	.2	.3	.4		.3	.3	.3	.3
.5	.3	.3	.3		.5	.4	.5	.4	.3		.7	.6	.5	.4	.3	.3	.4	.3	.6	.6	.6	.5	.3	.3		.8	.6	.5
.5	.3	.3	.3		.4	.5	.5	.5	.3		.5	.7	.5	.4	.4	.3	.4	.3	.6	.6	.5	.6	.4	.3	.8		.6	.5
.6	.4	.4	.4		.5	.5	.4	.5	.4		.4	.5	.5	.4	.5	.4	.5	.4	.6	.6	.5	.6	.5	.3	.6	.6		.7
.6	.4	.4	.4		.5	.5	.4	.4	.4		.4	.4	.4	.5	.5	.4	.5	.4	.6	.6	.5	.5	.5	.3	.5	.5	.7	



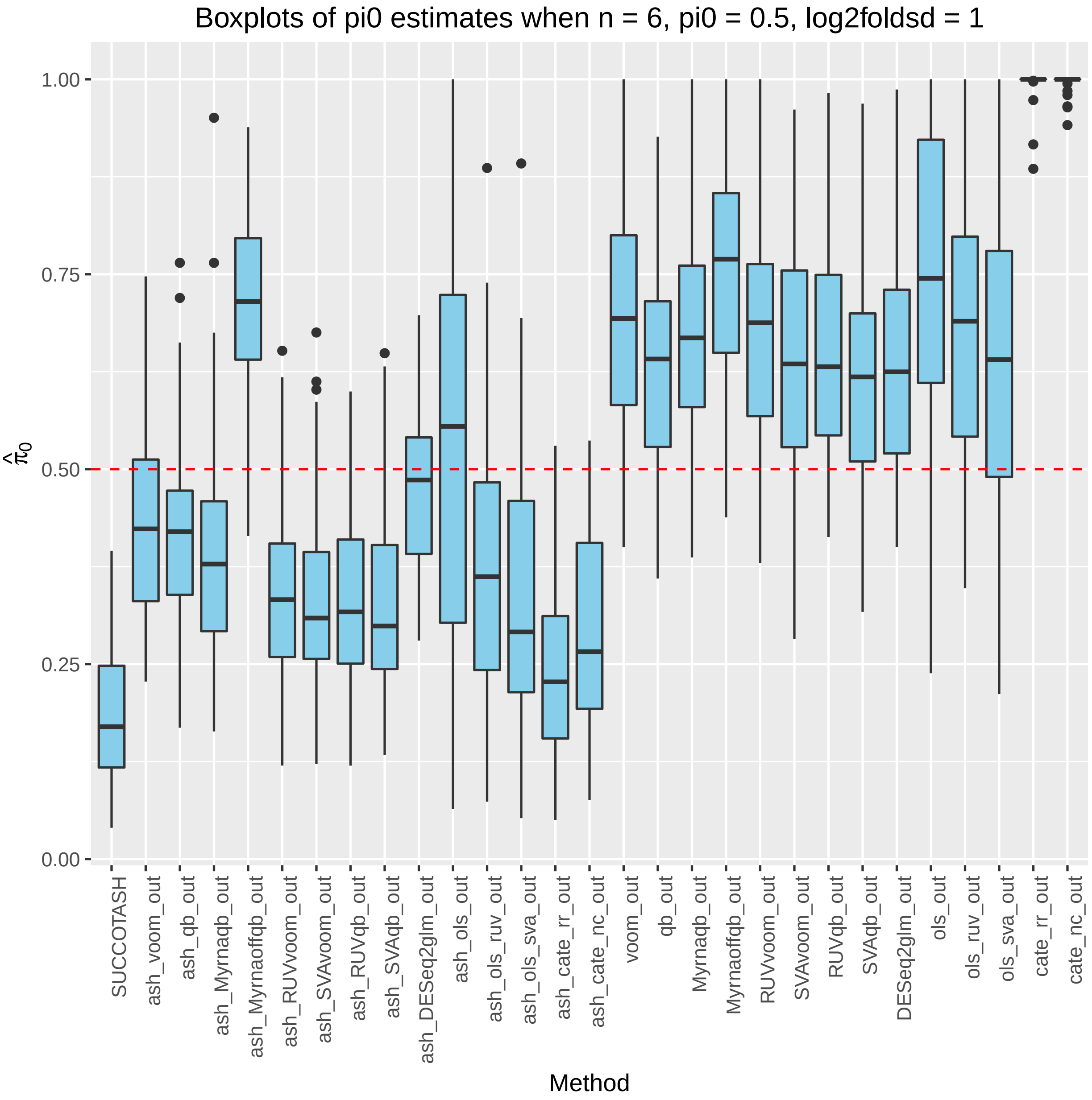
Boxplots of AUC when n = 10, pi0 = 0.9, log2foldsd = 1

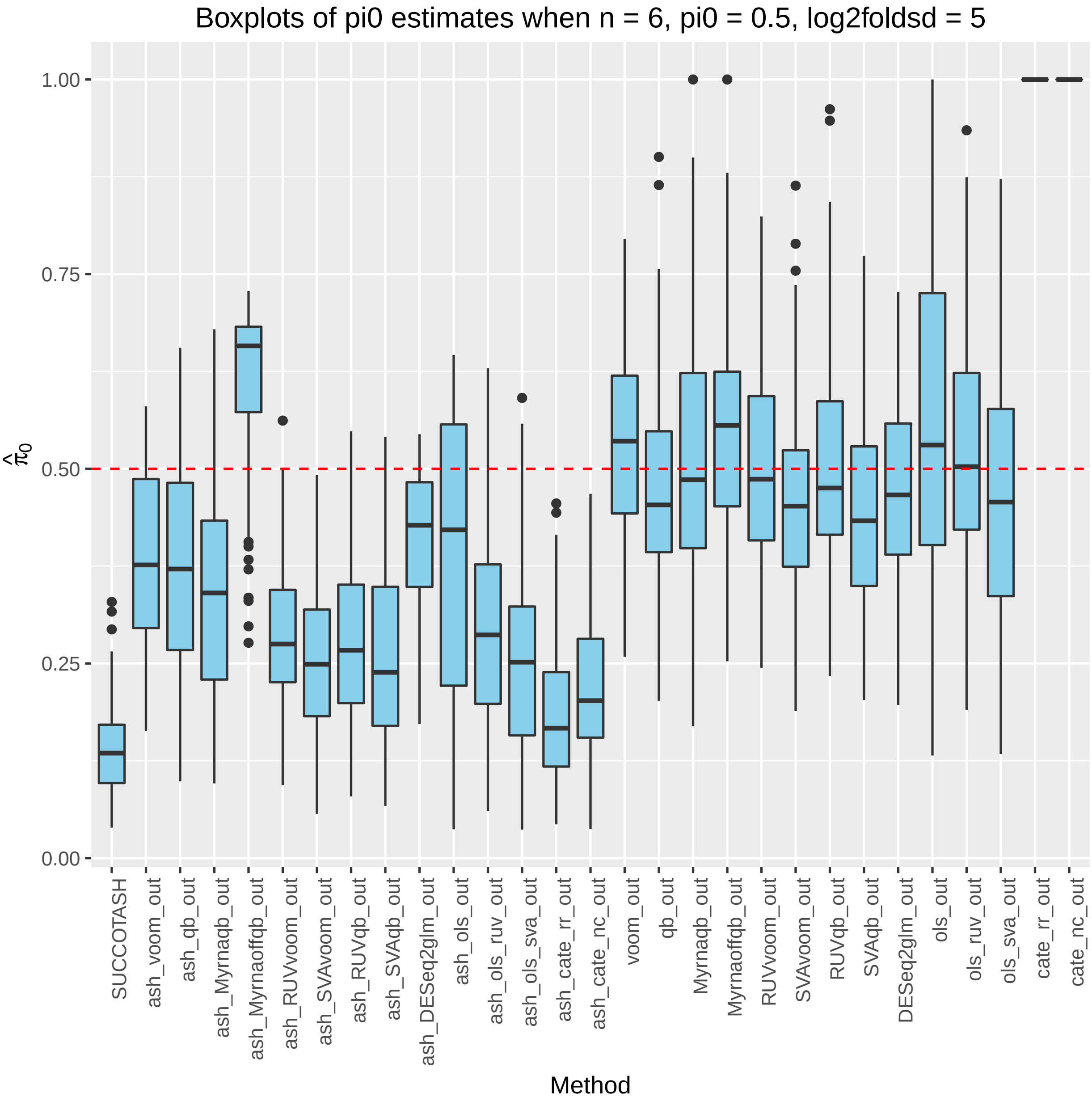


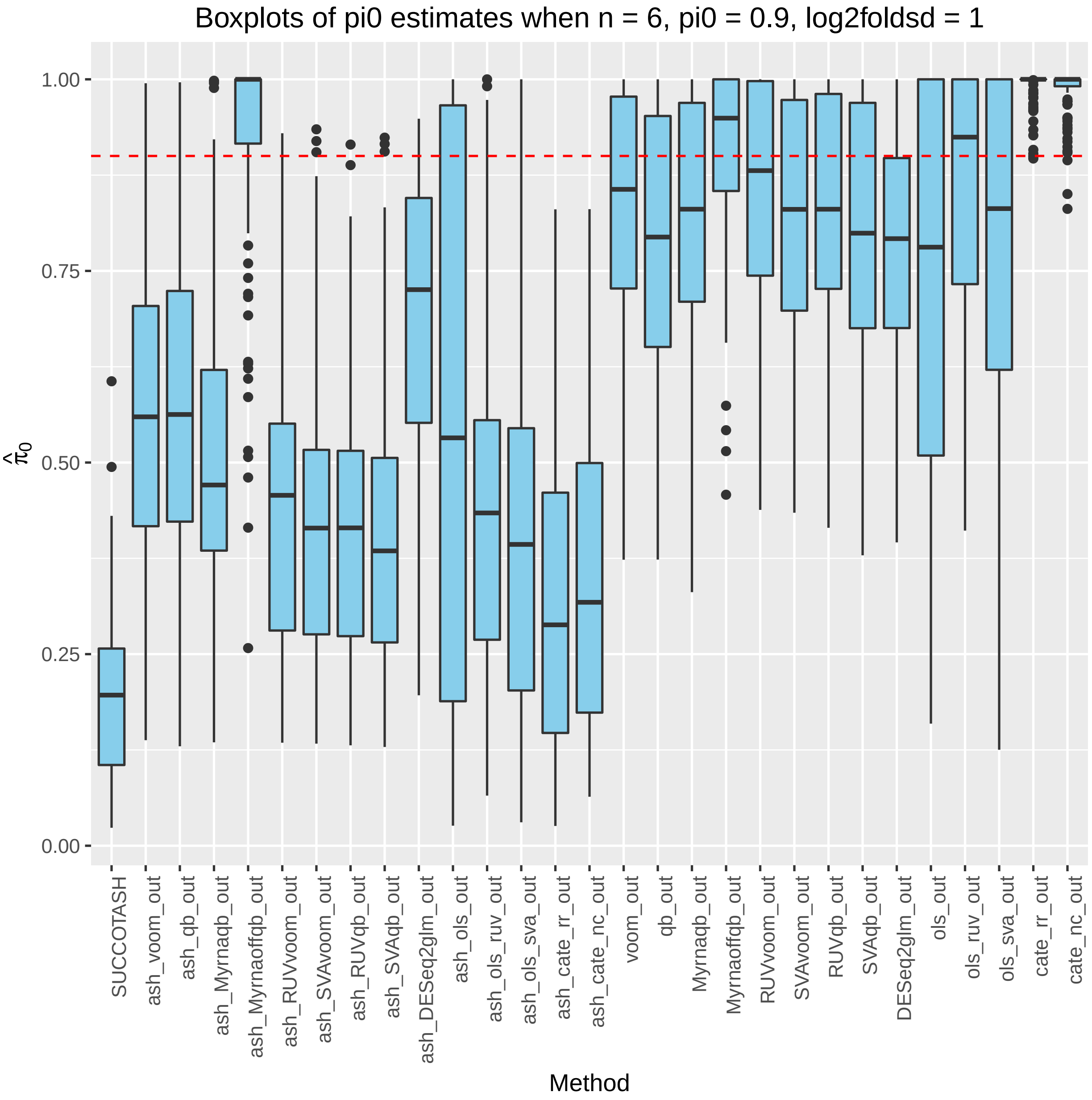
Median Kendall's Tau when n = 10 , pi0 = 0.9 , log2foldsd = 5

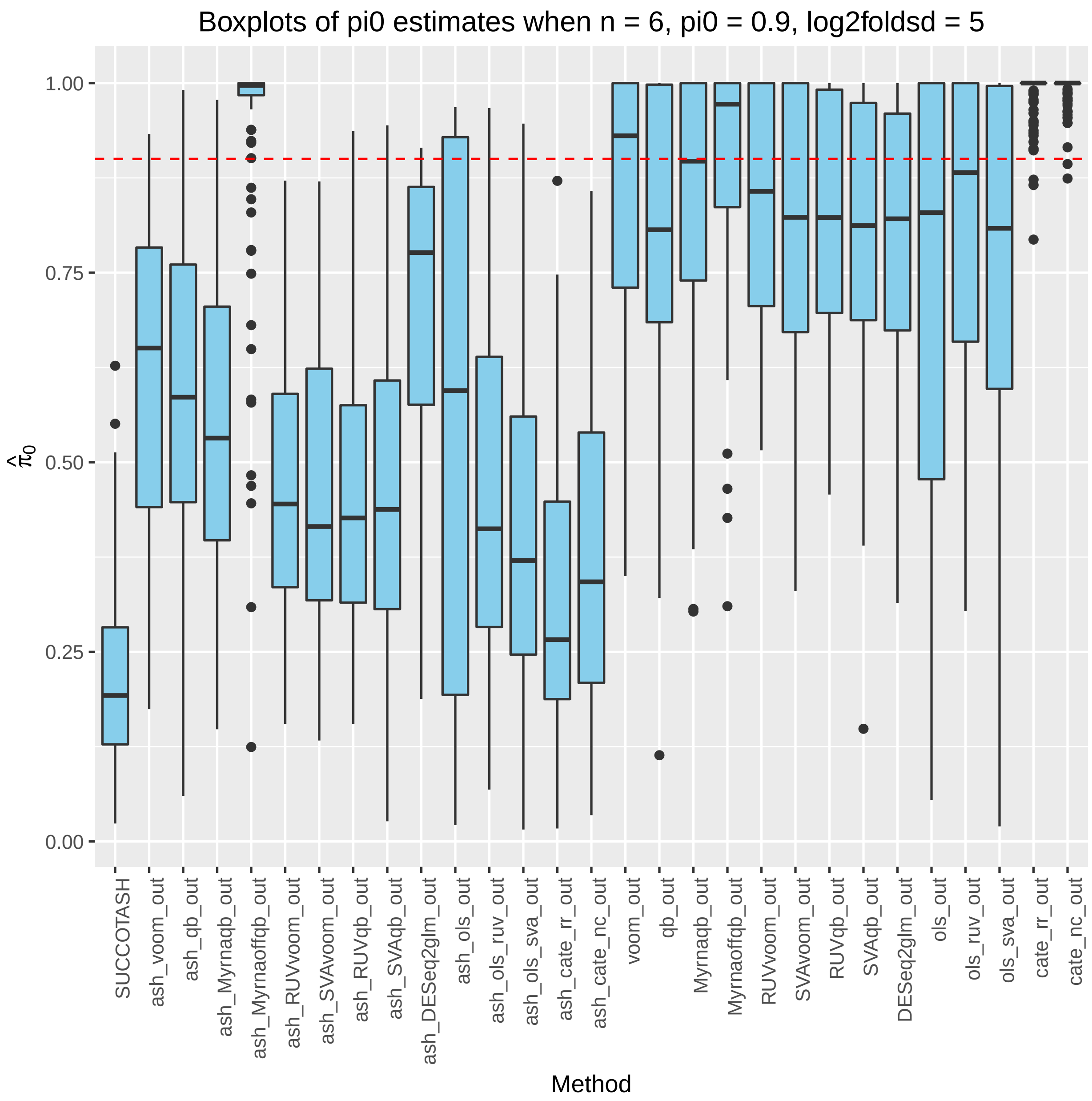


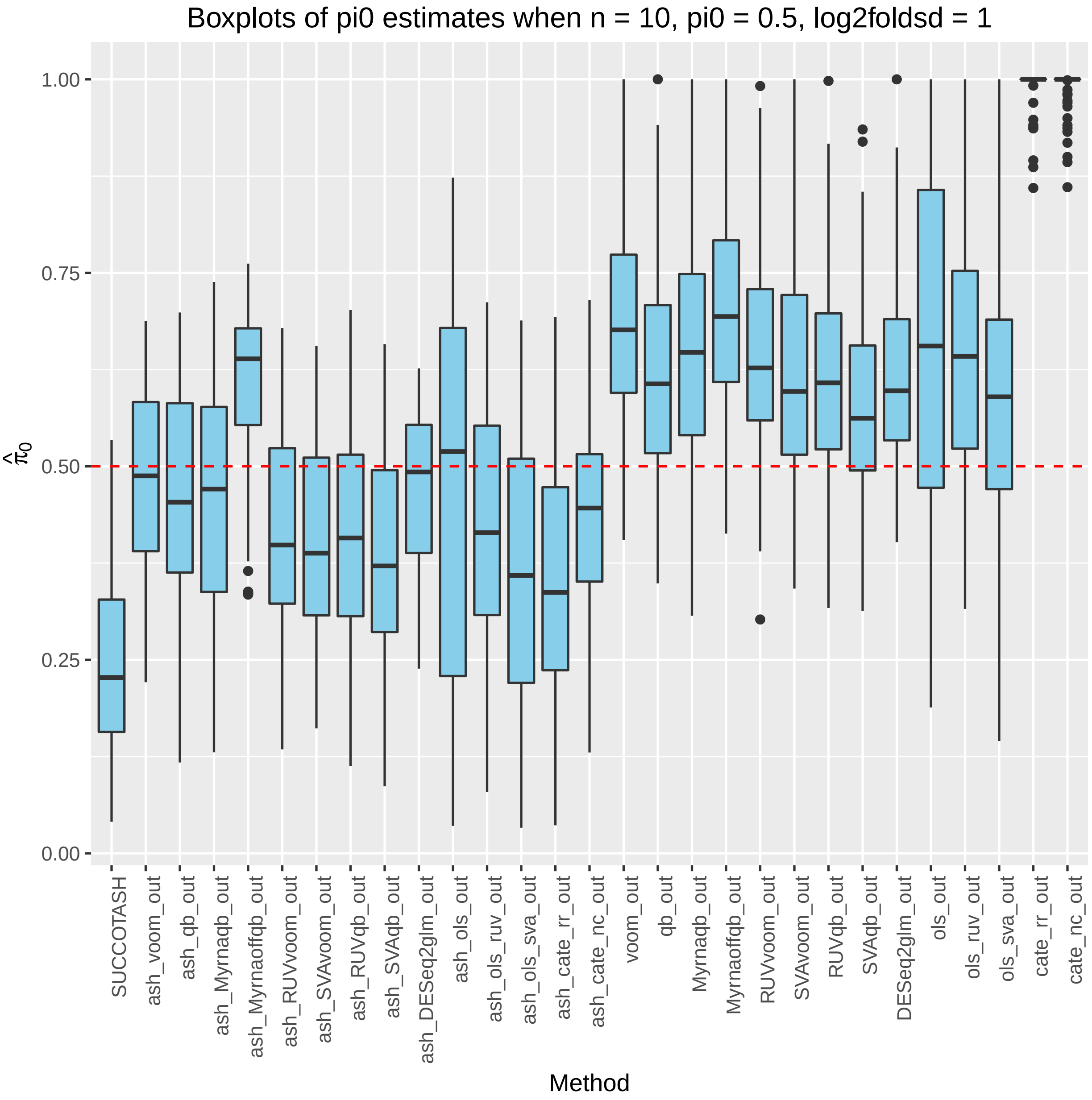
Boxplots of AUC when n = 10, pi0 = 0.9, log2foldsd = 5

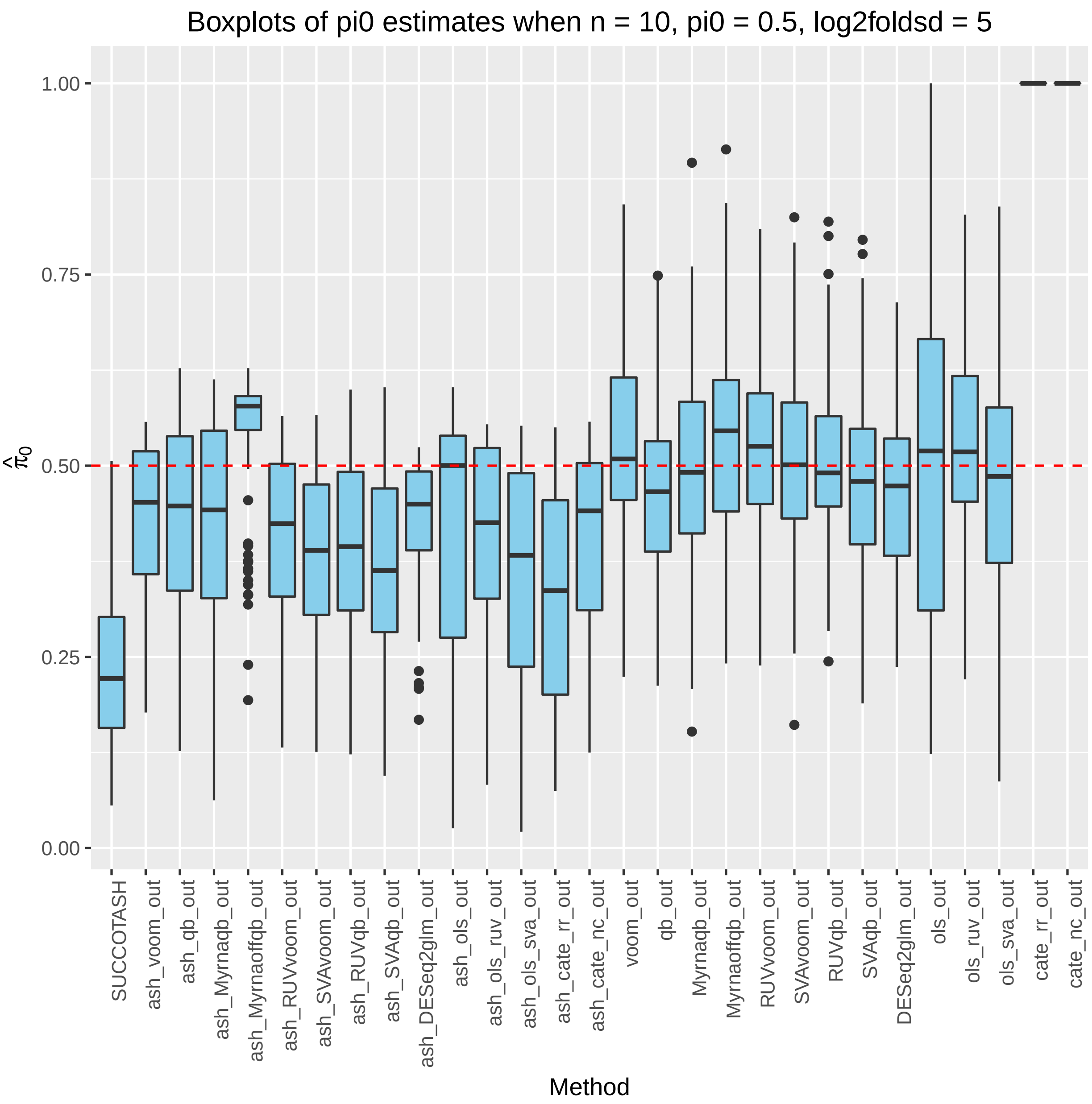

Boxplots of pi0 estimates when n = 6, pi0 = 0.5, log2foldsd = 1
$\hat{\pi}_0$
1.00
0.75
0.50
0.25
0.00
SUCCOTASH
ash_voom_out
ash_qb_out
ash_Myrnaqb_out
ash_Myrnaoffqb_out
ash_RUVvoom_out
ash_SVAvoom_out
ash_RUVqb_out
ash_SVAqb_out
ash_DESeq2glm_out
ash_ols_out
ash_ols_ruv_out
ash_ols_sva_out
ash_cate_rr_out
ash_cate_nc_out
voom_out
qb_out
Myrnaqb_out
Myrnaoffqb_out
RUVvoom_out
SVAvoom_out
RUVqb_out
SVAqb_out
DESeq2glm_out
ols_out
ols_ruv_out
ols_sva_out
cate_rr_out
cate_nc_out
Method

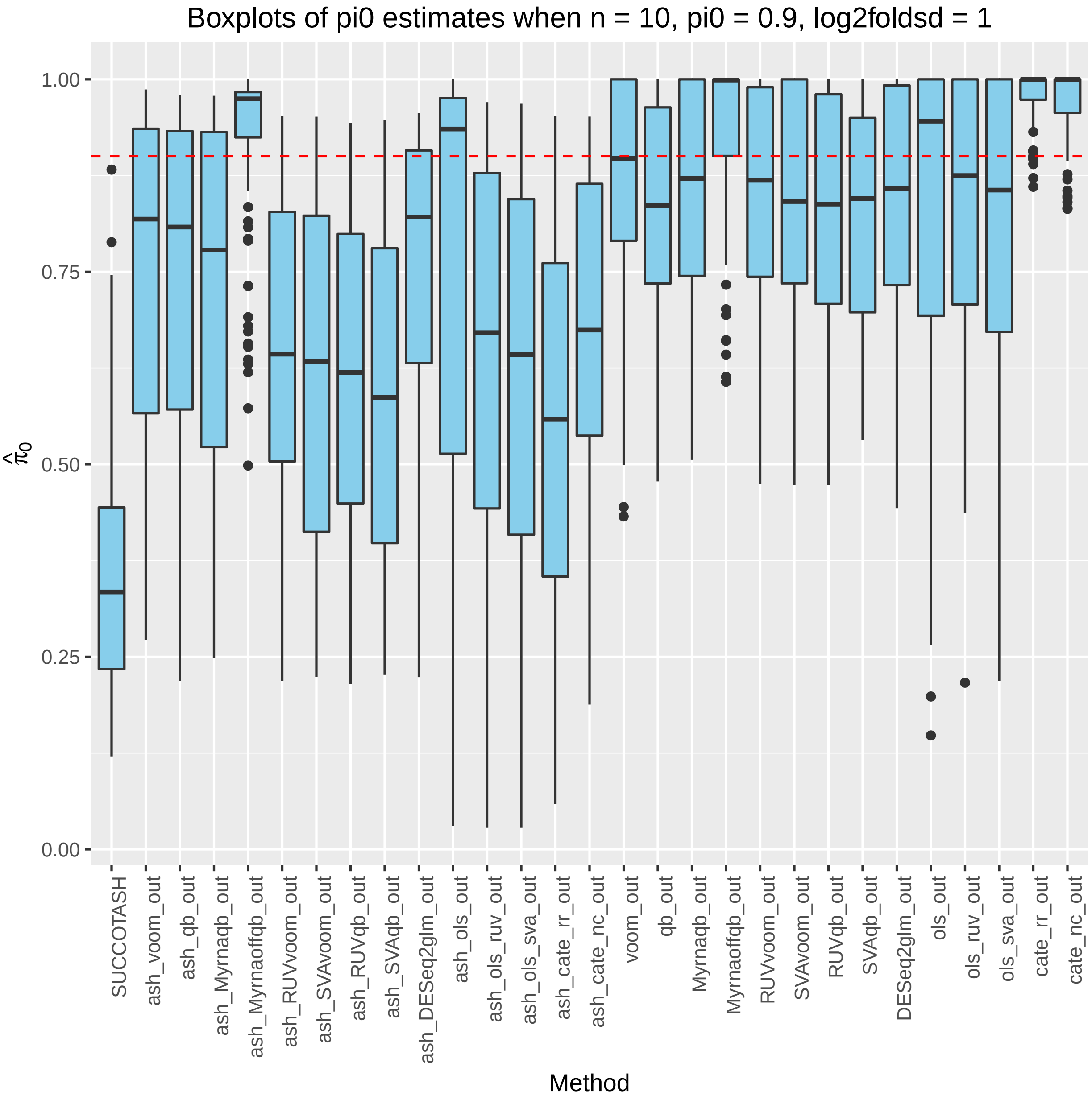

Boxplots of pi0 estimates when n = 6, pi0 = 0.5, log2foldsd = 5
$\hat{\pi}_0$
1.00
0.75
0.50
0.25
0.00
SUCCOTASH
ash_voom_out
ash_qb_out
ash_Myrnaqb_out
ash_Myrnaoffqb_out
ash_RUVvoom_out
ash_SVAvoom_out
ash_RUVqb_out
ash_SVAqb_out
ash_DESeq2glm_out
ash_ols_out
ash_ols_ruv_out
ash_ols_sva_out
ash_cate_rr_out
ash_cate_nc_out
voom_out
qb_out
Myrnaqb_out
Myrnaoffqb_out
RUVvoom_out
SVAvoom_out
RUVqb_out
SVAqb_out
DESeq2glm_out
ols_out
ols_ruv_out
ols_sva_out
cate_rr_out
cate_nc_out
Method

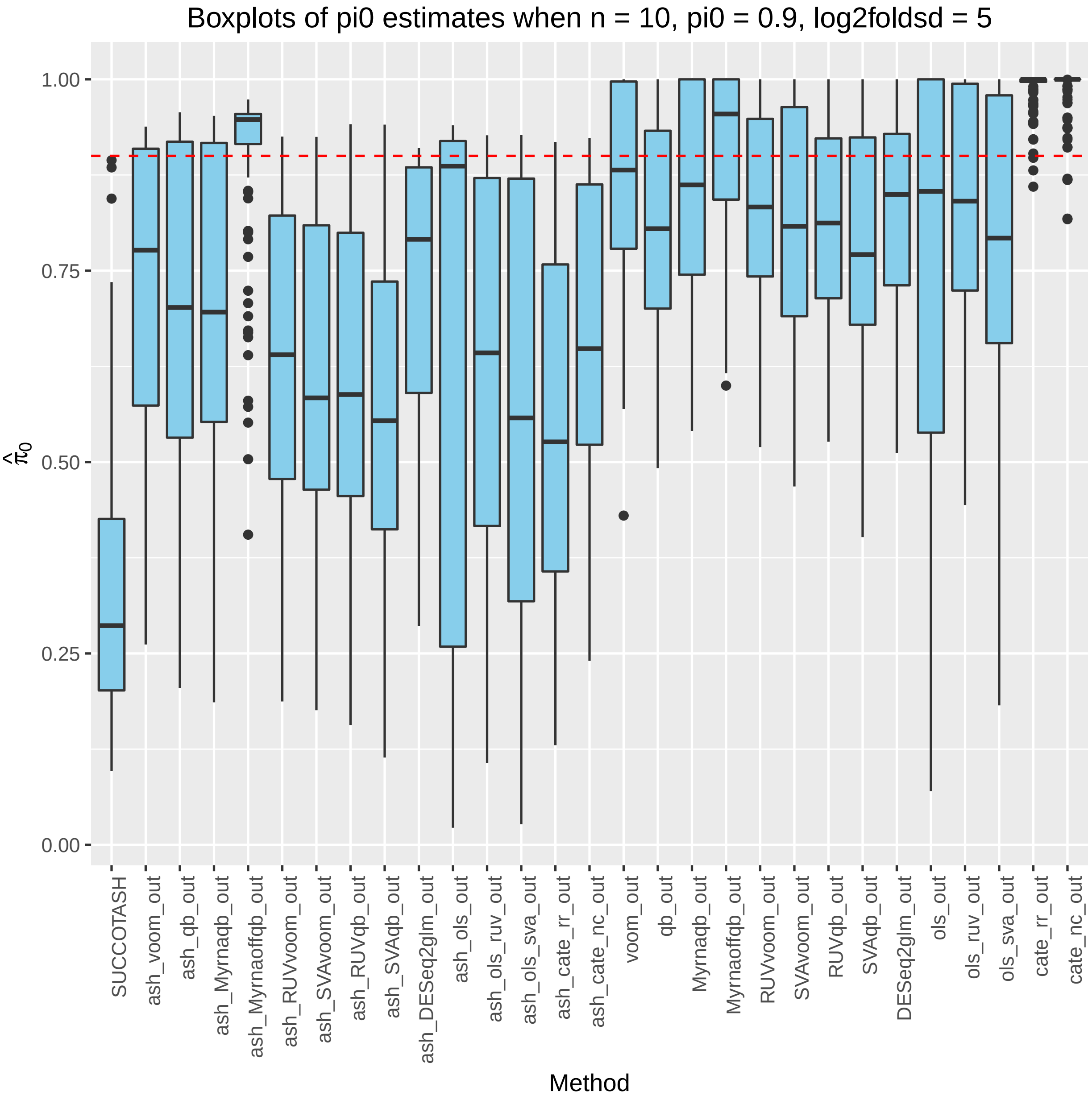

Boxplots of pi0 estimates when n = 6, pi0 = 0.9, log2foldsd = 1
$\hat{\pi}_0$
1.00
0.75
0.50
0.25
0.00
SUCCOTASH
ash_voom_out
ash_qb_out
ash_Myrnaqb_out
ash_Myrnaoffqb_out
ash_RUVvoom_out
ash_SVAvoom_out
ash_RUVqb_out
ash_SVAqb_out
ash_DESeq2glm_out
ash_ols_out
ash_ols_ruv_out
ash_ols_sva_out
ash_cate_rr_out
ash_cate_nc_out
voom_out
qb_out
Myrnaqb_out
Myrnaoffqb_out
RUVvoom_out
SVAvoom_out
RUVqb_out
SVAqb_out
DESeq2glm_out
ols_out
ols_ruv_out
ols_sva_out
cate_rr_out
cate_nc_out
Method


Boxplots of pi0 estimates when n = 6, pi0 = 0.9, log2foldsd = 5
$\hat{\pi}_0$
1.00
0.75
0.50
0.25
0.00
SUCCOTASH
ash_voom_out
ash_qb_out
ash_Myrnaqb_out
ash_Myrnaoffqb_out
ash_RUVvoom_out
ash_SVAvoom_out
ash_RUVqb_out
ash_SVAqb_out
ash_DESeq2glm_out
ash_ols_out
ash_ols_ruv_out
ash_ols_sva_out
ash_cate_rr_out
ash_cate_nc_out
voom_out
qb_out
Myrnaqb_out
Myrnaoffqb_out
RUVvoom_out
SVAvoom_out
RUVqb_out
SVAqb_out
DESeq2glm_out
ols_out
ols_ruv_out
ols_sva_out
cate_rr_out
cate_nc_out
Method


Boxplots of pi0 estimates when n = 10, pi0 = 0.5, log2foldsd = 1
$\hat{\pi}_0$
1.00
0.75
0.50
0.25
0.00
SUCCOTASH
ash_voom_out
ash_qb_out
ash_Myrnaqb_out
ash_Myrnaoffqb_out
ash_RUVvoom_out
ash_SVAvoom_out
ash_RUVqb_out
ash_SVAqb_out
ash_DESeq2glm_out
ash_ols_out
ash_ols_ruv_out
ash_ols_sva_out
ash_cate_rr_out
ash_cate_nc_out
voom_out
qb_out
Myrnaqb_out
Myrnaoffqb_out
RUVvoom_out
SVAvoom_out
RUVqb_out
SVAqb_out
DESeq2glm_out
ols_out
ols_ruv_out
ols_sva_out
cate_rr_out
cate_nc_out
Method


Boxplots of pi0 estimates when n = 10, pi0 = 0.5, log2foldsd = 5
$\hat{\pi}_0$
1.00
0.75
0.50
0.25
0.00
SUCCOTASH
ash_voom_out
ash_qb_out
ash_Myrnaqb_out
ash_Myrnaoffqb_out
ash_RUVvoom_out
ash_SVAvoom_out
ash_RUVqb_out
ash_SVAqb_out
ash_DESeq2glm_out
ash_ols_out
ash_ols_ruv_out
ash_ols_sva_out
ash_cate_rr_out
ash_cate_nc_out
voom_out
qb_out
Myrnaqb_out
Myrnaoffqb_out
RUVvoom_out
SVAvoom_out
RUVqb_out
SVAqb_out
DESeq2glm_out
ols_out
ols_ruv_out
ols_sva_out
cate_rr_out
cate_nc_out
Method


Boxplots of pi0 estimates when n = 10, pi0 = 0.9, log2foldsd = 1
$\hat{\pi}_0$
1.00
0.75
0.50
0.25
0.00
SUCCOTASH
ash_voom_out
ash_qb_out
ash_Myrnaqb_out
ash_Myrnaoffqb_out
ash_RUVvoom_out
ash_SVAvoom_out
ash_RUVqb_out
ash_SVAqb_out
ash_DESeq2glm_out
ash_ols_out
ash_ols_ruv_out
ash_ols_sva_out
ash_cate_rr_out
ash_cate_nc_out
voom_out
qb_out
Myrnaqb_out
Myrnaoffqb_out
RUVvoom_out
SVAvoom_out
RUVqb_out
SVAqb_out
DESeq2glm_out
ols_out
ols_ruv_out
ols_sva_out
cate_rr_out
cate_nc_out
Method



References

Jushan Bai, Kunpeng Li, et al. Statistical analysis of factor models of high dimension. *The Annals of Statistics*, 40(1):436–465, 2012.

Andreas Buja and Nermin Eyuboglu. Remarks on parallel analysis. *Multivariate behavioral research*, 27(4):509–540, 1992.

J Gagnon-Bartsch, L Jacob, and TP Speed. Removing unwanted variation from high dimensional data with negative controls. Technical report, Technical Report 820, Department of Statistics, University of California, Berkeley, 2013.

Ben Langmead, Kasper D Hansen, Jeffrey T Leek, et al. Cloud-scale rna-sequencing differential expression analysis with myrna. *Genome Biol*, 11(8):R83, 2010.

Charity W Law, Yunshun Chen, Wei Shi, and Gordon K Smyth. Voom: precision weights unlock linear model analysis tools for rna-seq read counts. *Genome Biol*, 15(2):R29, 2014.

Jeffrey T Leek. svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic acids research*, page gku864, 2014.

Jeffrey T Leek and John D Storey. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet*, 3(9):1724–1735, 2007.

Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biol*, 15(12):550, 2014.

Davide Risso, John Ngai, Terence P Speed, and Sandrine Dudoit. Normalization of rna-seq data using factor analysis of control genes or samples. *Nature biotechnology*, 32(9):896–902, 2014.

Gordon K Smyth. Limma: linear models for microarray data. In *Bioinformatics and computational biology solutions using R and Bioconductor*, pages 397–420. Springer, 2005.

John D Storey. A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):479–498, 2002.

Yunting Sun, Nancy R Zhang, Art B Owen, et al. Multiple hypothesis testing adjusted for latent variables, with an application to the agemap gene expression data. *The Annals of Applied Statistics*, 6(4):1664–1688, 2012.

Jingshu Wang, Qingyuan Zhao, Trevor Hastie, and Art B Owen. Confounder adjustment in multiple hypotheses testing. *arXiv preprint arXiv:1508.04178*, 2015.