Report_2

Guanshengrui Hao

12/10/2018

Investigate issues in pooling permuted test statistics

The last report compared the performance of two multiple hypothesis strategies, individual strategy and pooling strategy. Based on the results last time, we can see that the pooling strategy performs significantly better (higher statistical power) than the individual strategy under cases when

- the null distribution is continuously distributed;
- the number of samples in each group is small.

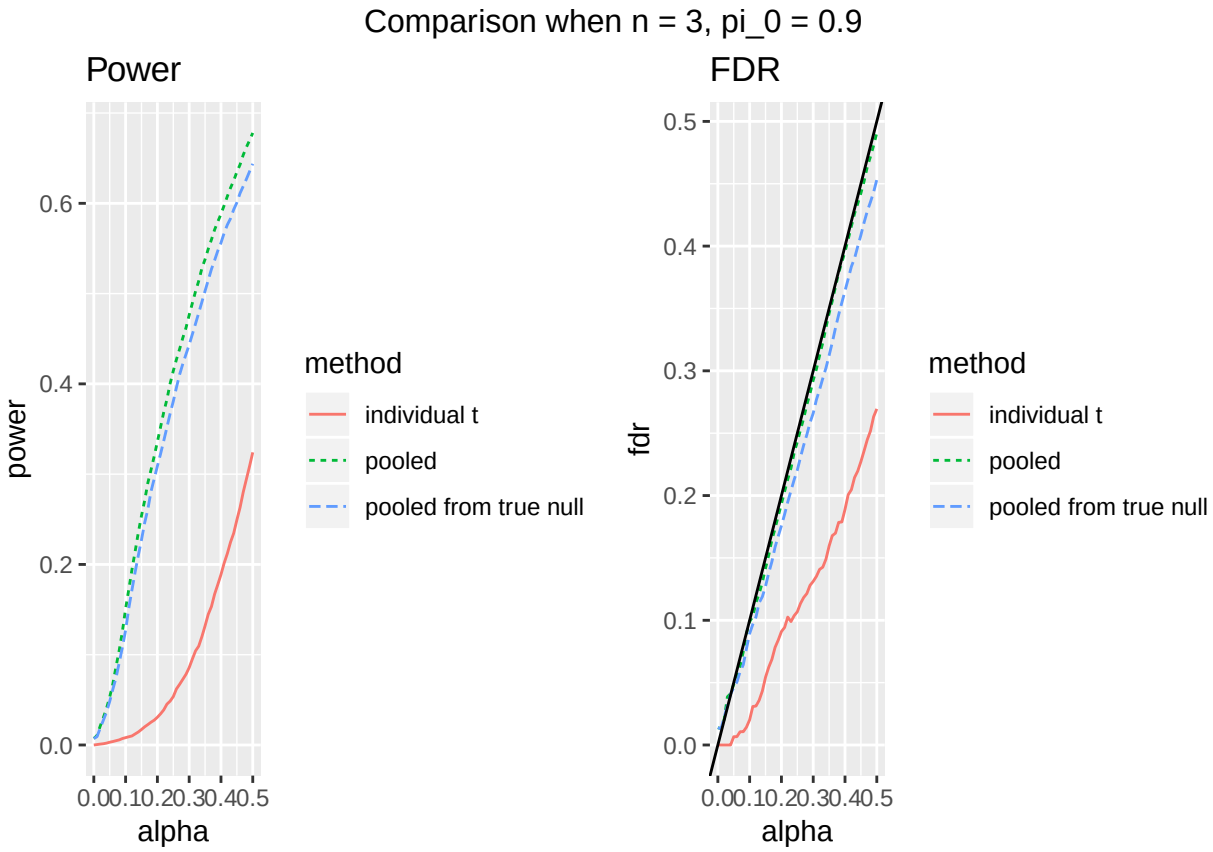
Although showing very promising performance, the pooling strategy often has realized FDR higher than the nominal level, especially under the setups where the true null proportion π_0 is close to 1 and the true null hypothesis has heavy tails (Cauchy case in last report). I investigated the issue and found three possible causes.

1. The current pooling strategy is not perfect.
   The current pooling stragety pools over t-statistics from both true nulls and true non-nulls. Under permutation, the t-statistics from true non-nulls tends to be more concentrated that those from true nulls (recall the t-statistic under true non-null from the balanced permutation). The reference distribution constructed this way would provide more extreme p-values, thus causing the beyond-controlled FDR. The ideal pooling stragety should be pooling test statistics under only true null from all possible permutations.
2. The reference distribution cannot cover the whole support.
   When the value of a test statistic from the original grouping goes beyond the support of the reference distribution, its p-value would be given as 0, meaning it would definitely be rejected not matter what adjustment is applied. If such a test statistic happens to be from true null, then the FDR would be positive even if the nominal level is 0. Such problem can be reflected by the positive intercept term when regressing the realized FDR against the nominal level α. Furthermore, this problem is more severe when the null distribution is heavy-tailed, as it's more likely to have test statistics go beyond the support of reference distribution.
3. The choice of test statistic used is not ideal.
   When the true null distribution is normal, it seems to be fine after addressing the first two causes. However, when the true null distribution is Cauchy, the slope is still higher than it is supposed to be. This might be caused by the choice of the test statistic.

I will illustrate those causes using simulations under different setups.

Under homogeneous normal setup

In this comparison, all samples under true null are drawn from $N(0, 1)$ distribution. Under true non-null, samples from control group are drawn from $N(0, 1)$ while samples from treatment group are drawn from $N(\mu, \sigma^2)$, where $\mu \sim \text{Unif}(1, 4)$, $\sigma \sim \text{Exp}(1)$. The multiple testing procedure repeats 50 times.



We can see from the plots that although simply pooling all permuted test statistics can offer a larger statistical power, its FDR can get beyond the nominal level. To compare with, if we only pool test statistics from the true null cases (assuming we know the true nulls, as in the simulation), we could control the FDR under nominal level (under the normality setup). More evidence by regressing realized FDR against nominal level α.

```
### Pooling all test statistics
summary(lm(fdr_pool_mean~alpha))
```

```
##
## Call:
## lm(formula = fdr_pool_mean ~ alpha)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.006032 -0.002464 -0.001049  0.002484  0.015414
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.001525   0.001085  -1.405    0.166
## alpha        0.986267   0.003741 263.630   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.003933 on 49 degrees of freedom
## Multiple R-squared:  0.9993, Adjusted R-squared:  0.9993
## F-statistic: 6.95e+04 on 1 and 49 DF,  p-value: < 2.2e-16
```

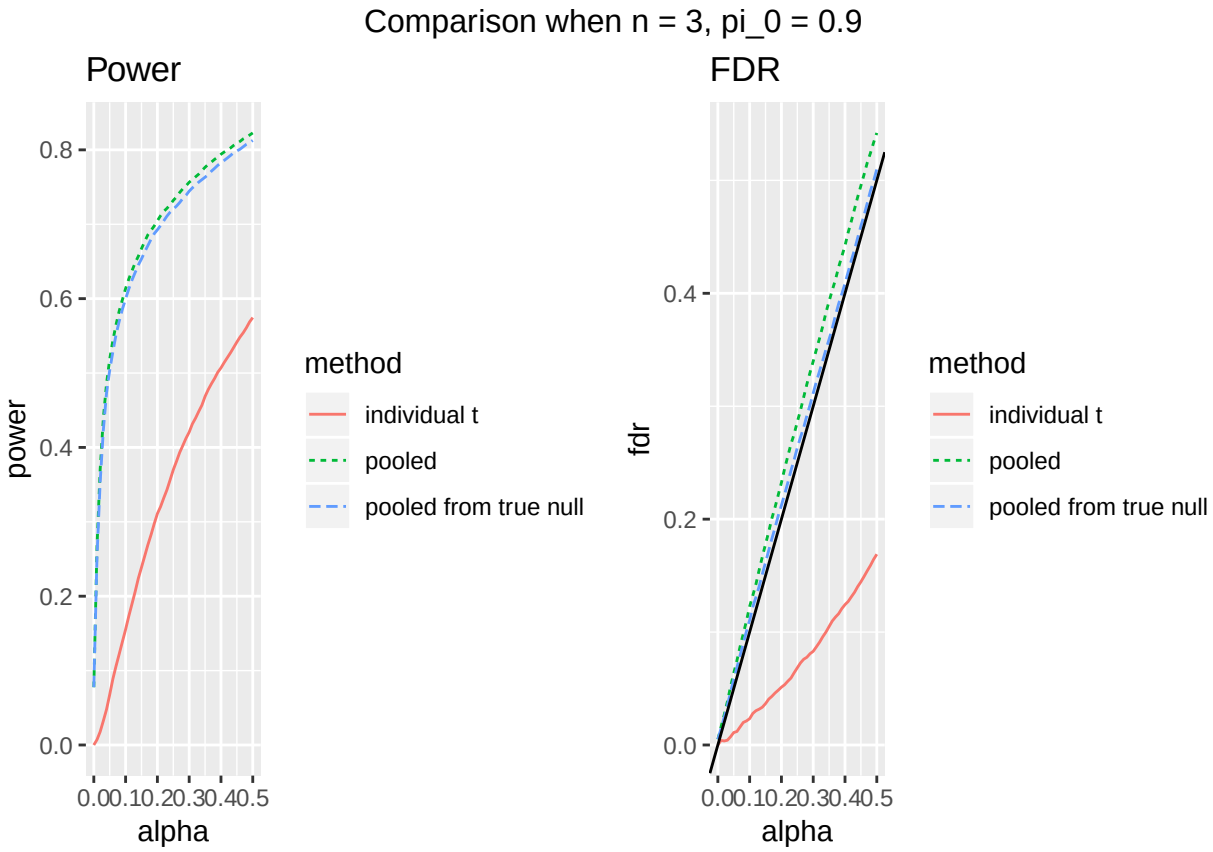
```
### Pooling only test statistics under true nulls
summary(lm(fdr_pool_ideal_mean~alpha))
```

```
##
## Call:
## lm(formula = fdr_pool_ideal_mean ~ alpha)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.006292 -0.002872 -0.001793  0.002935  0.016474
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.002585   0.001198  -2.157   0.0359 *
## alpha        0.907763   0.004131 219.743   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.004342 on 49 degrees of freedom
## Multiple R-squared:  0.999,  Adjusted R-squared:  0.999
## F-statistic: 4.829e+04 on 1 and 49 DF,  p-value: < 2.2e-16
```

As we are using BH procedure to control FDR, it is supposed to be controlled below $\pi_0\alpha$ level. We can see the slope of the first regression (pooling all) goes beyond 0.9, while the second is controlled below 0.9.

Under homogeneous cauchy setup

In this comparison, all samples under true null are drawn from Cauch$(0, 1)$. Under true non-null, samples from control group are drawn from Cauch$(0, 1)$, while samples from treatment group are drawn from Cauch(l, s), where $l \sim \text{Unif}(5, 10)$, $s \sim \text{Exp}(1)$. The multiple testing procedure repeats 50 times.



Regressing realized FDR against nominal level α for both cases (pooling all and pooling only true null),

```
### Pooling all test statistics
summary(lm(fdr_pool_mean~alpha))
```

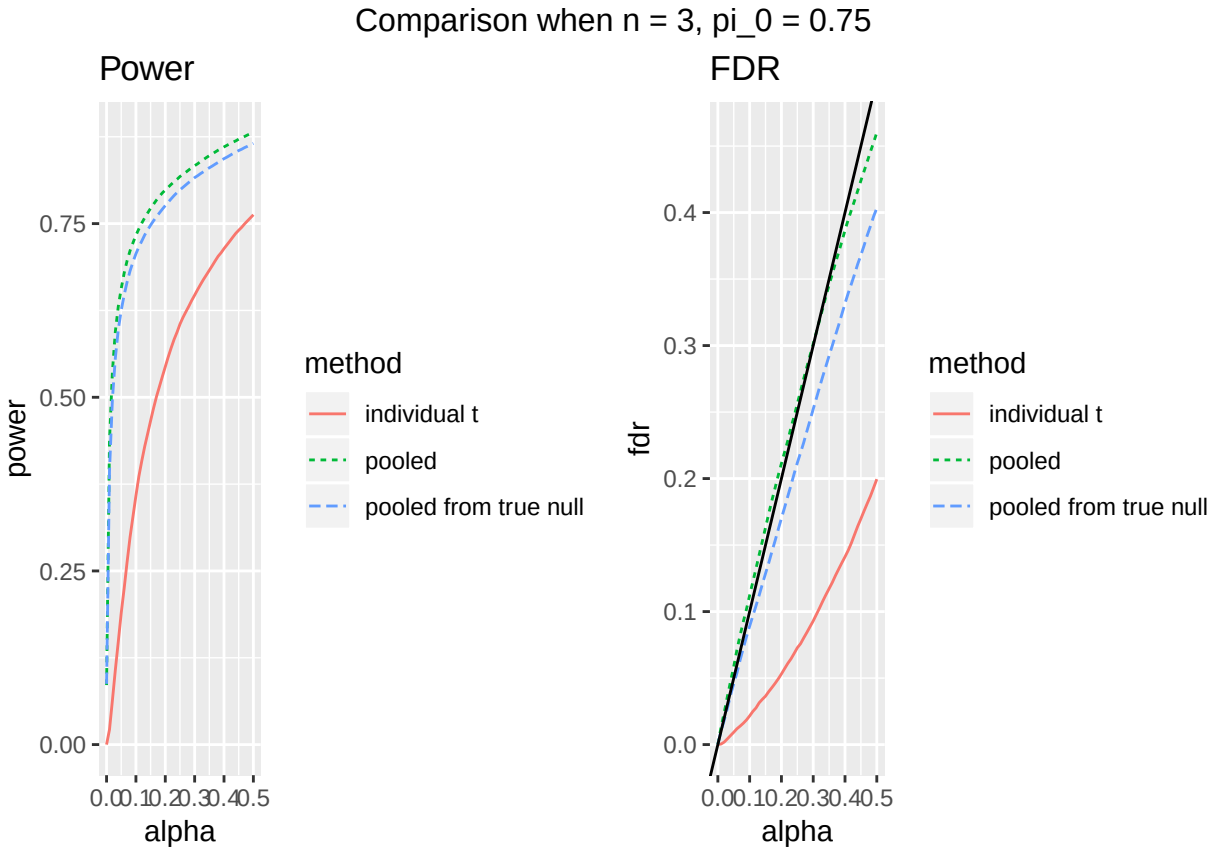
```
##
## Call:
## lm(formula = fdr_pool_mean ~ alpha)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0098297 -0.0022901  0.0007157  0.0034023  0.0054883
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.013269   0.001154    11.5  1.6e-15 ***
## alpha       1.077188   0.003978   270.8  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.004182 on 49 degrees of freedom
## Multiple R-squared:  0.9993, Adjusted R-squared:  0.9993
## F-statistic: 7.332e+04 on 1 and 49 DF,  p-value: < 2.2e-16
```

```
### Pooling only test statistics under true nulls
summary(lm(fdr_pool_ideal_mean~alpha))
```

```
##
```

```
## Call:
## lm(formula = fdr_pool_ideal_mean ~ alpha)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0050072 -0.0013275  0.0001477  0.0016398  0.0031404
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0094856  0.0005861   16.18   <2e-16 ***
## alpha       1.0030142  0.0020201  496.51   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.002124 on 49 degrees of freedom
## Multiple R-squared:  0.9998, Adjusted R-squared:  0.9998
## F-statistic: 2.465e+05 on 1 and 49 DF,  p-value: < 2.2e-16
```

we can see the intercepts are both significantly positive (as explained in the second cause) and the slopes are both larger than what it is supposed to be ($\pi_0 = 0.9$).



Regressing realized FDR against nominal level α for both cases (pooling all and pooling only true null),

```
### Pooling all test statistics
summary(lm(fdr_pool_mean~alpha))
```

```
##
## Call:
```

```
## lm(formula = fdr_pool_mean ~ alpha)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.020261 -0.005132  0.002396  0.006989  0.008451
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.020380   0.002304   8.845    1e-11 ***
## alpha       0.913782   0.007942 115.054   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.008349 on 49 degrees of freedom
## Multiple R-squared:  0.9963, Adjusted R-squared:  0.9962
## F-statistic: 1.324e+04 on 1 and 49 DF,  p-value: < 2.2e-16
```

```
### Pooling only test statistics under true nulls
summary(lm(fdr_pool_ideal_mean~alpha))
```

```
##
## Call:
## lm(formula = fdr_pool_ideal_mean ~ alpha)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.007749 -0.001421  0.000904  0.001944  0.003143
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0069419  0.0007839   8.856 9.67e-12 ***
## alpha       0.8083672  0.0027019 299.185  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.00284 on 49 degrees of freedom
## Multiple R-squared:  0.9995, Adjusted R-squared:  0.9994
## F-statistic: 8.951e+04 on 1 and 49 DF,  p-value: < 2.2e-16
```

we can see the intercepts are both significantly positive (as explained in the second cause) and the slopes are both larger than what it is supposed to be ($\pi_0 = 0.75$).