

```

[1] "# of discriminatory genes = 5"

[1] "fold change for discriminatory genes: 0.3"

[1] "fold change for nondiscriminatory genes: 0"

choose_k
  2
100

```

The optimal number of clusters K is set as the most frequently found K from the 100 simulations tabulated above. It is found by using the BIC criterion after running the unpenalized EM algorithm on K spanning from 2 to 8.

Using the last set of simulated counts, I ran a grid search across varying tuning parameters. As done in Pan et al, I fixed  $\lambda_1 = 1$ , and searched over  $\lambda_2 = (0.1, 0.2, \dots, 2)$  and  $\tau = (0.1, 0.2, \dots, 2)$ :

```

[1] "tau= 0.2" "tau= 0.3" "tau= 0.1"

[1] "lambda1= 1"

[1] "lambda2= 0.2" "lambda2= 0.3" "lambda2= 0.4"

```

The results of the final run based on optimal tuning parameters are below:  
Below are the summary of results:

```

[1] "Mean pi: 0.630869565217391" "Mean pi: 0.369130434782609"

[1] "First 3 genes:"

      [,1]      [,2]
[1,] 3.562222 4.152810
[2,] 4.799893 5.387755
[3,] 4.308123 4.901663

[1] "Last 3 genes:"

      [,1]      [,2]
[98,] 6.397815 6.398028
[99,] 5.952756 5.950407
[100,] 6.377911 6.379097

[1] "Mean % of nondiscriminatory genes: 0.9499"

[1] "Mean ARI: 1"

[1] "Mean sensitivity: 0.998"

[1] "Mean false positive rate: 0.000210526315789474"

```