

Purity and likelihood plots vs alpha

2020-07-11

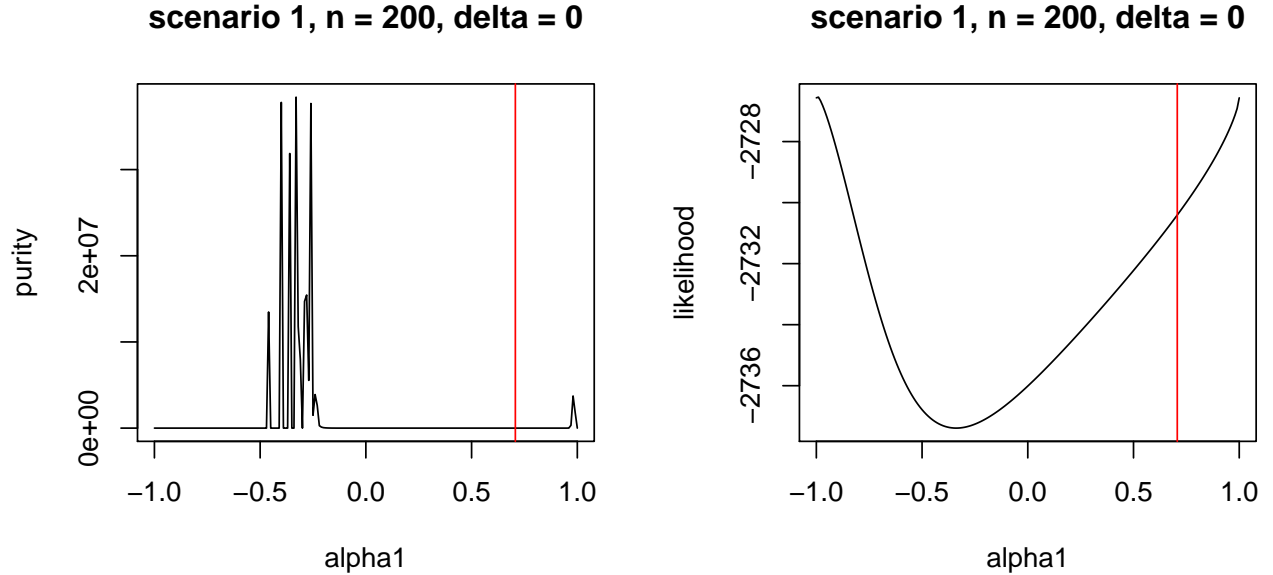
Parameter setting 1

In the last week's simulation, the purity vs alpha plot has weird shape. It might be because that the elements are small in the covariance matrix, i.e.,

- $D_{drg} = D_{pbo} = \begin{pmatrix} 0.1 & 0.01 & 0.01 \\ 0.01 & 0.1 & 0.01 \\ 0.01 & 0.01 & 0.05 \end{pmatrix}$
- $n = 200$
- dimension of the predictors $p = 2$
- $\beta_{drg} = \beta_{pbo} = (1, -0.05, -0.02)$
- $\Gamma_{drg} = (0, -\sin(\frac{\pi}{3})/10, -\cos(\frac{\pi}{3})/10)$
- $\Gamma_{pbo} = (0, \cos(\frac{\pi}{3})/10, -\sin(\frac{\pi}{3})/10)$
- $S = [1, t, t^2]$, $t = [0, 1, 2, 3, 4, 6, 8]$ is the design matrix for fixed effect and random effect
- $x \sim MVN(\mu_x, \Sigma_x)$, $\mu_x = \mathbf{0}_p$, Σ_x has diagonal equals to 1 and 0.5 everywhere else.
- $\epsilon_{drg}, \epsilon_{pbo} \sim N(0, 1^2)$

Suppose we have two predictors, true $\alpha_0 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. For different $\alpha = (x_1, \sqrt{1-x_1^2})$, $x_1 \in (-1, 1)$, the purity and log-likelihood are calculated.

The figures of purity and log-likelihood vs x_1 are:



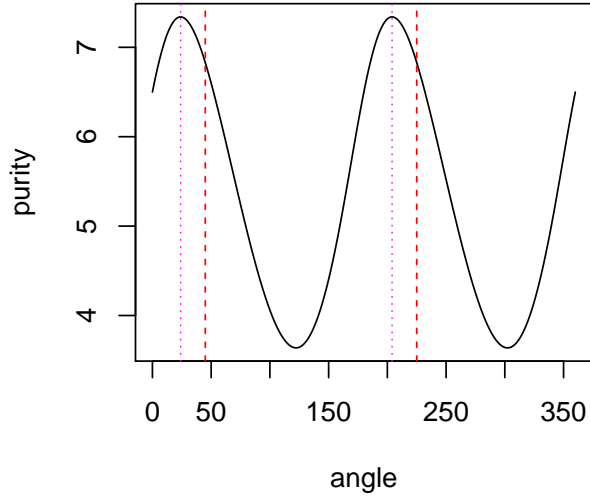
Parameter setting 2

If we change the parameters, the plots can be much better.

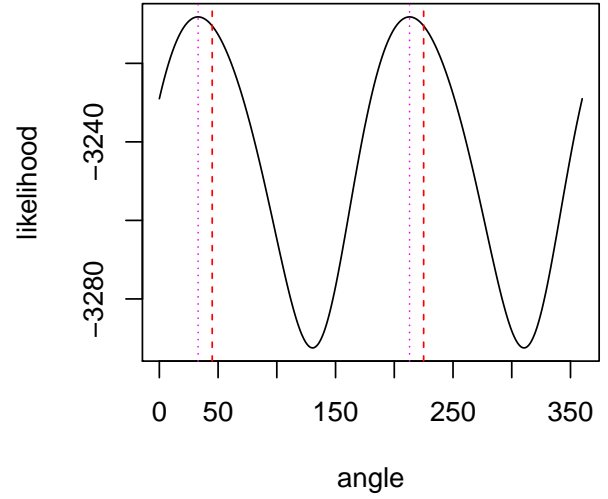
Let

- $n = 200, p = 2, \alpha_0 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}) = (\cos(\frac{\pi}{4}), \sin(\frac{\pi}{4}))$
- $\beta_{drg} = \beta_{pbo} = (1, -0.05, -0.02)'$
- $\Gamma_{drg} = (0, -\sin(\frac{\pi}{3}), -\cos(\frac{\pi}{3}))$
- $\Gamma_{pbo} = (0, \cos(\frac{\pi}{3}), -\sin(\frac{\pi}{3}))$
- $S = [1, t, t^2], t = [0, 1, 2, 3, 4, 6, 8]$ is the design matrix for fixed effect and random effect
- $x \sim MVN(\mu_x, \Sigma_x), \mu_x = \mathbf{0}_p, \Sigma_x$ has diagonal equals to 1 and 0.5 everywhere else.
- $D_{drg} = D_{pbo} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$
- $\epsilon_{drg}, \epsilon_{pbo} \sim N(0, 1^2)$

scenario 2, n = 200, delta = 0

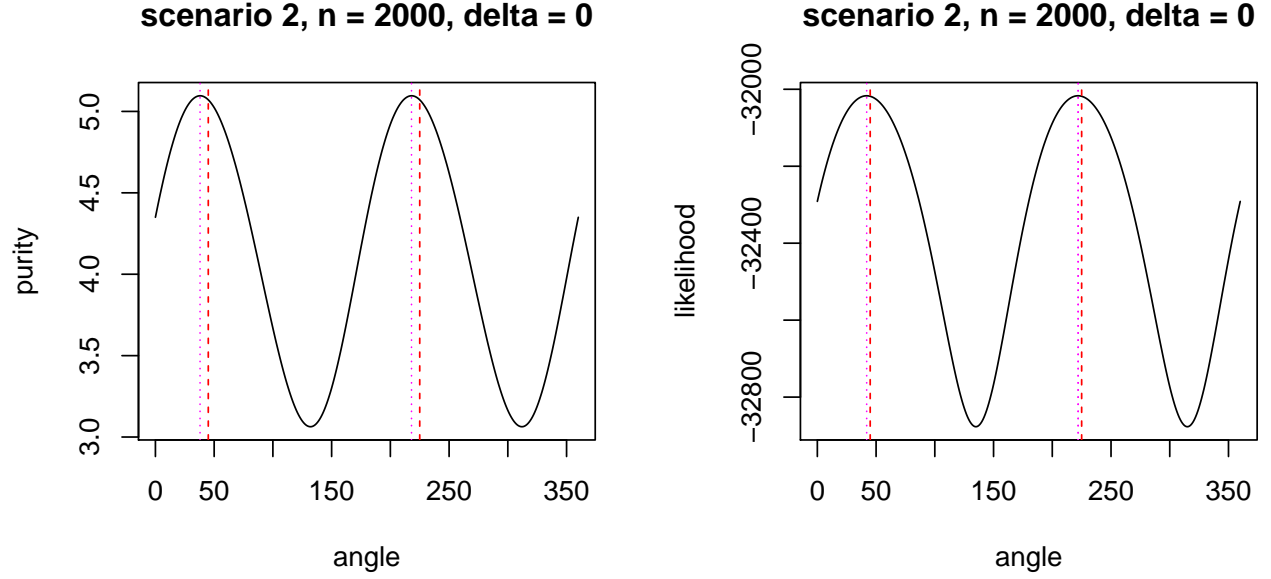


scenario 2, n = 200, delta = 0



where the dish red line shows the position of true α and the purple dot line shows the α that maximizes the purity or log likelihood.

If we increase the sample size, the estimation can be more accurate and the difference between estimated α and true α are smaller



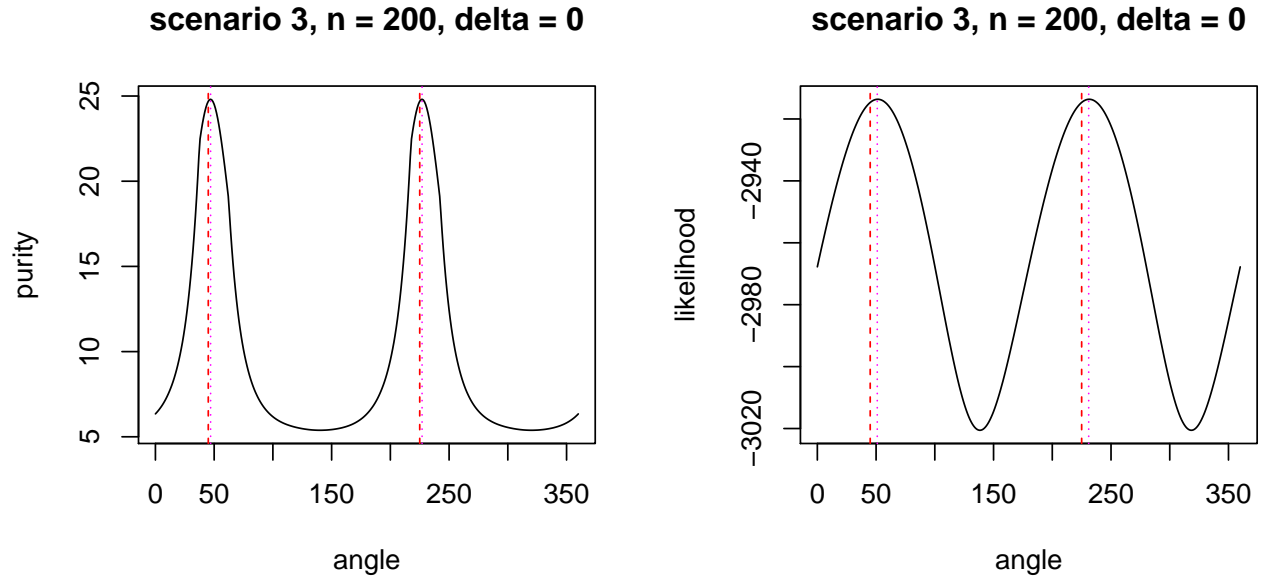
Therefore, how well is the α estimated depends on the parameter setting.

Parameter setting 3

If we consider another covariance matrix

$$\bullet D_{drg} = \begin{pmatrix} 1.45 & -0.11 & 0.2 \\ -0.11 & 0.17 & -0.08 \\ 0.2 & -0.08 & 0.23 \end{pmatrix}, D_{pbo} = \begin{pmatrix} 1.03 & -0.23 & -0.15 \\ -0.23 & 0.68 & 0.25 \\ -0.15 & 0.25 & 1.36 \end{pmatrix}$$

The plots look like:



That is, with small sample size, in this parameter setting, both of the methods can give good estimation.

When drug and placebo group have different α

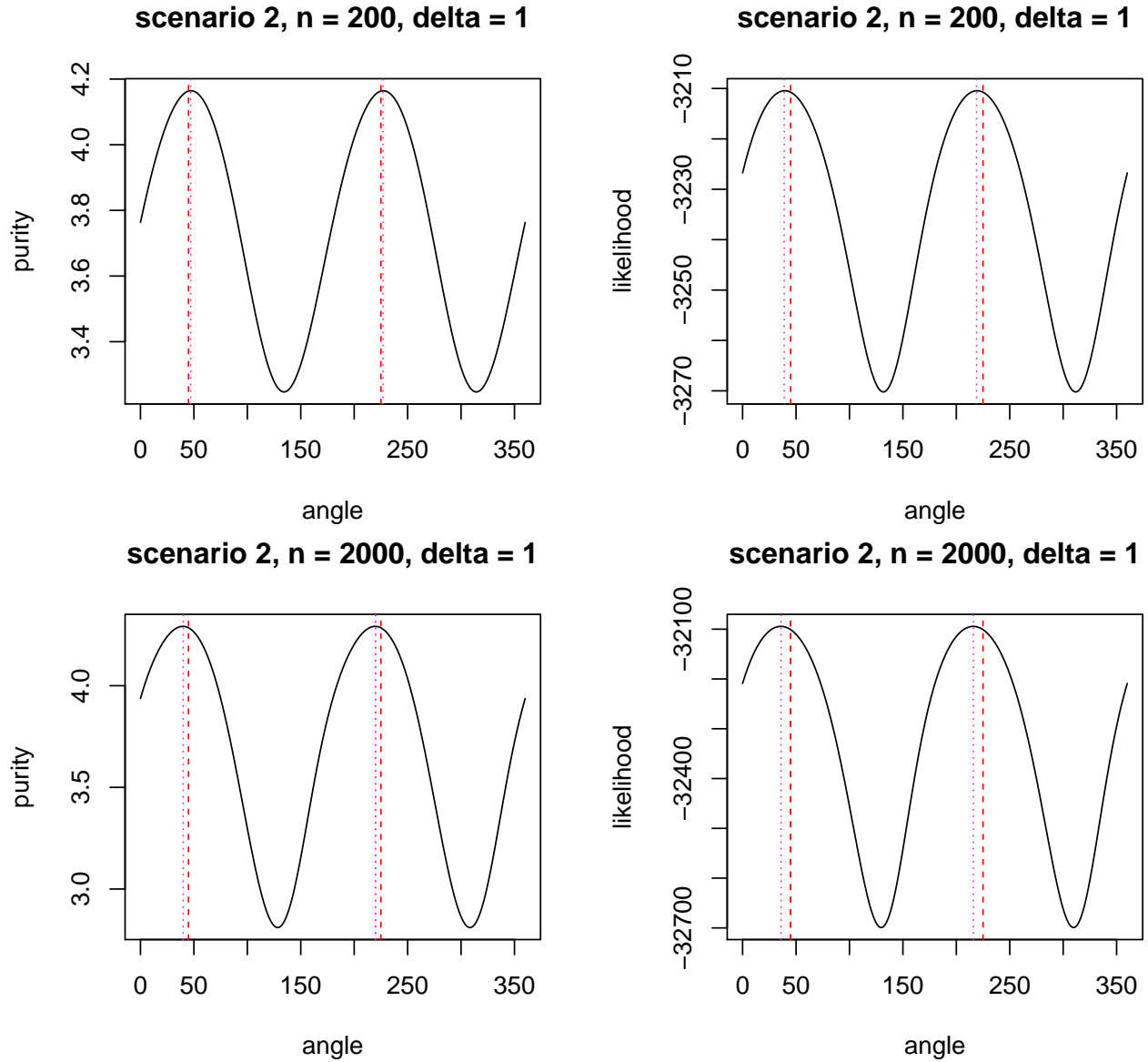
If we still have the previous parameter settings for data generation, while the true α are different from drug and placebo group,

- $\alpha = \alpha_1 + \delta\alpha_{2k}, k = 1, 2,$
 - $\alpha_1 = (1, 1)$
 - drg: $\alpha_{21} = (1, 2)$
 - pbo: $\alpha_{22} = (-2, -1)$

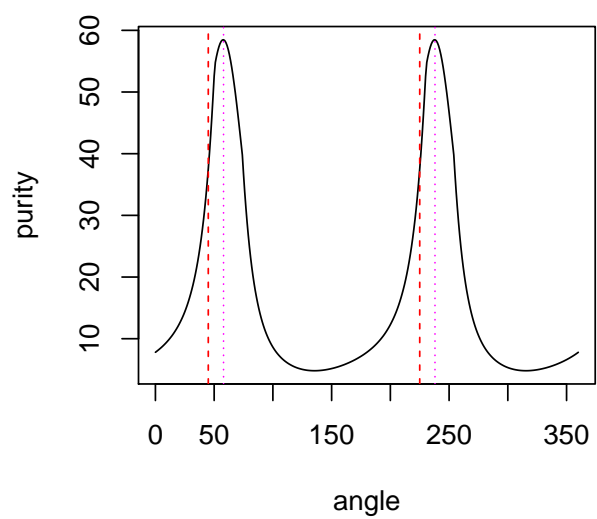
If $\delta = 0$, $\alpha_{drg} = \alpha_{pbo} = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. The cosine similarity is 1.

If $\delta = 1$, cosine similarity is -0.55

The plots for thoes scenarios are



scenario 3, n = 200, delta = 1



scenario 3, n = 200, delta = 1

