

# Simulation to estimating alpha 2

2019-09-11

## Kullback-Leibler divergence

For distributions  $F_1$  and  $F_2$  of a continuous random variable, the KL divergence is defined as:

$$D_{KL}(F_1||F_2) = \int_{-\infty}^{+\infty} f_1(x) \log\left(\frac{f_1(x)}{f_2(x)}\right) dx \quad (1)$$

where  $f_1$  and  $f_2$  denote the probability density of  $F_1$  and  $F_2$ .

Therefore, we defined the *purity* as

$$\begin{aligned} & \int \log(f_1)f_1 - \int \log(f_2)f_1 + \int \log(f_2)f_2 - \int \log(f_1)f_2 \\ &= \left( -\frac{p}{2} \log(2\pi) - \frac{1}{2} \log(|D_1|) - \frac{p}{2} \right) \\ & - \left( -\frac{p}{2} \log(2\pi) - \frac{1}{2} \log(|D_2|) - \frac{1}{2} \left( \text{tr}(D_2^{-1}D_1) + (\mu_1 - \mu_2)' D_2^{-1} (\mu_1 - \mu_2) \right) \right) \\ &+ \left( -\frac{p}{2} \log(2\pi) - \frac{1}{2} \log(|D_2|) - \frac{p}{2} \right) \\ & - \left( -\frac{p}{2} \log(2\pi) - \frac{1}{2} \log(|D_1|) - \frac{1}{2} \left( \text{tr}(D_1^{-1}D_2) + (\mu_1 - \mu_2)' D_1^{-1} (\mu_1 - \mu_2) \right) \right) \\ &= -p + \frac{1}{2} \text{tr}(D_2^{-1}D_1) + \frac{1}{2} \text{tr}(D_1^{-1}D_2) + \frac{1}{2} (\mu_1 - \mu_2)' (D_1^{-1} + D_2^{-1}) (\mu_1 - \mu_2) \end{aligned}$$

where  $\mu_1 = \beta_1 + \Gamma_1 \alpha' x$ ,  $\mu_2 = \beta_2 + \Gamma_2 \alpha' x$ .

That is:

$$\text{KL purity} = -p + \frac{1}{2} \text{tr}(D_2^{-1}D_1) + \frac{1}{2} \text{tr}(D_1^{-1}D_2) + \frac{1}{2} (\mu_1 - \mu_2)' (D_1^{-1} + D_2^{-1}) (\mu_1 - \mu_2)$$

The optimization of  $\alpha$ :

$$\alpha = -2B^{-1}A'$$

where

- $A = \sum_i (\Gamma_1 - \Gamma_2)' (D_1^{-1} + D_2^{-1}) (\beta_1 - \beta_2) x_i'$
- $B = \sum_i x_i \left( (\Gamma_1 - \Gamma_2)' (D_1^{-1} + D_2^{-1}) (\Gamma_1 - \Gamma_2) \right) x_i'$

## Simulation Setting

Sample size:

- 200 subjects in each group, 400 in total.

Dimensions:

- $p = 4$ . (4 baseline biosignature)

Angle between  $\Gamma_1$  and  $\Gamma_3$ :

- scenario: 0, 30, 60, 90, 120, 150, 180 degrees.

Initial value: True  $\alpha$

Simulation time: 500

Parameters

- Each subject has 7 time points:  $2 \times 100 \times 7 = 1400$
- True  $\alpha = [0.5, 0.5, 0.5, 0.5]'$
- $\beta_{pbo} = [0, 3, 0.9]'$ ,  $\beta_{drg} = [0, 3.1, 1]'$
- $\Gamma_{drg} = [0, 1, 0]'$ , angle between the two  $\Gamma$  lines can be 0, 30, 60, 90, 120, 150, and 180 degrees.
- $|\Gamma_{drg}| = |\Gamma_{pbo}| = 1$

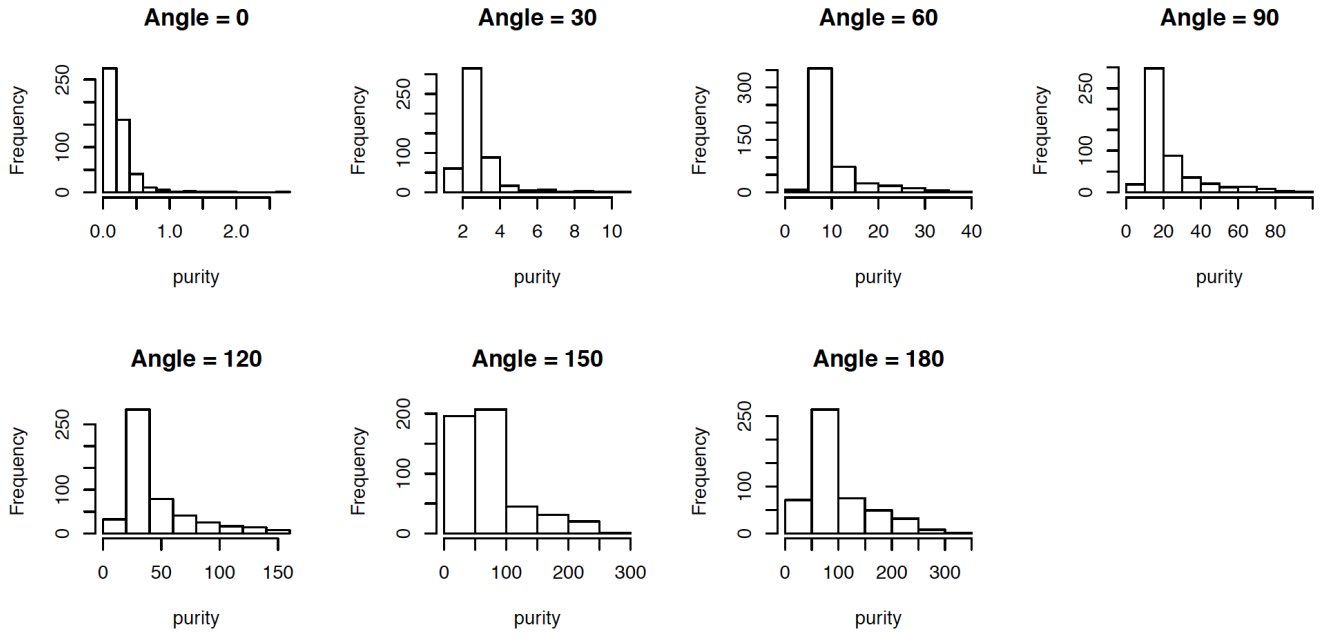
## Result

$\alpha$  calculated by  $\hat{\beta}_{drg}, \hat{\beta}_{pbo}, \hat{\Gamma}_{drg}, \hat{\Gamma}_{pbo}, \hat{D}_{drg}, \hat{D}_{pbo}$

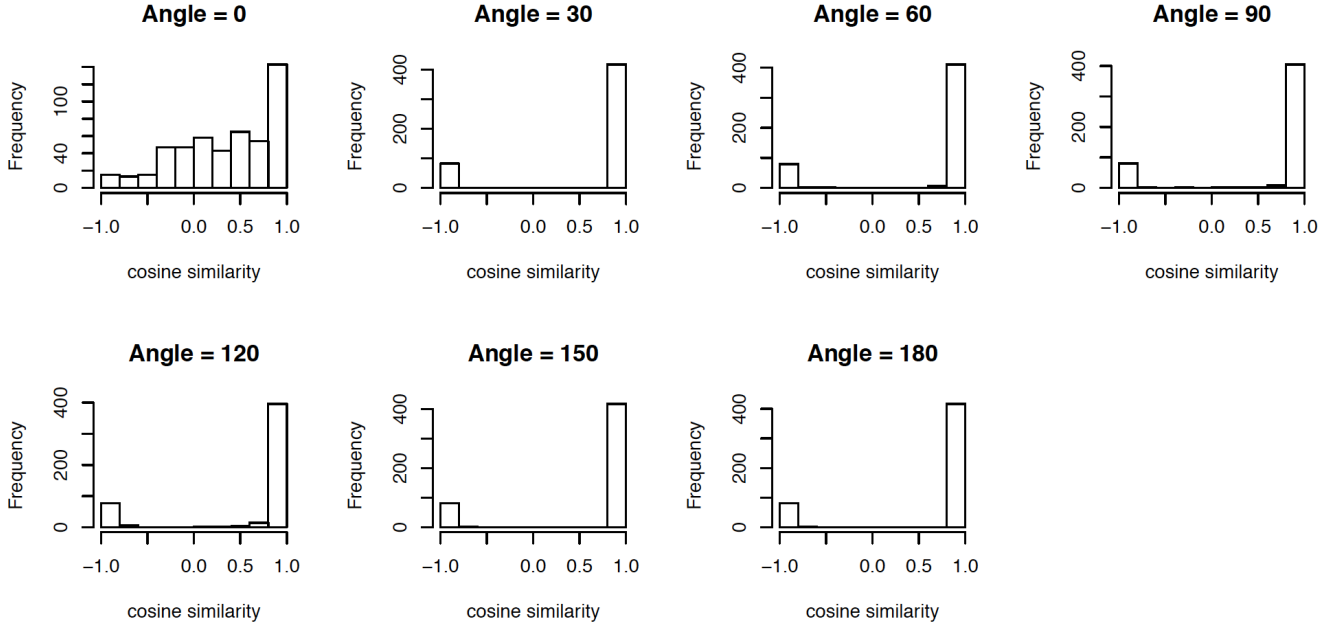
Angle	Mean KL dif	Mean KL	SD KL	Mean cosine	SD cosine
0	0.24	0.24	0.24	0.35	0.54
30	0.70	2.79	1.07	0.66	0.73
60	3.37	10.18	5.64	0.66	0.72
90	9.47	22.76	15.26	0.65	0.73
120	21.22	44.90	30.80	0.64	0.72
150	35.49	75.20	49.32	0.67	0.73
180	44.41	96.89	57.38	0.66	0.74

- Mean\_KL\_dif: the mean absolute differences between true *purity* and the estimated  $\hat{purity}$  during the 500 times replication.
- Mean\_KL: The mean value of the estimated *purity*
- SD\_KL: The standard deviation value of the estimated *purity*
- Mean\_cosine: the mean value of the cosine\_similarity.
- SD\_cosine: The standard deviation value of the cosine\_similarity

**The histogram of purity**



The histogram of the consine similarity



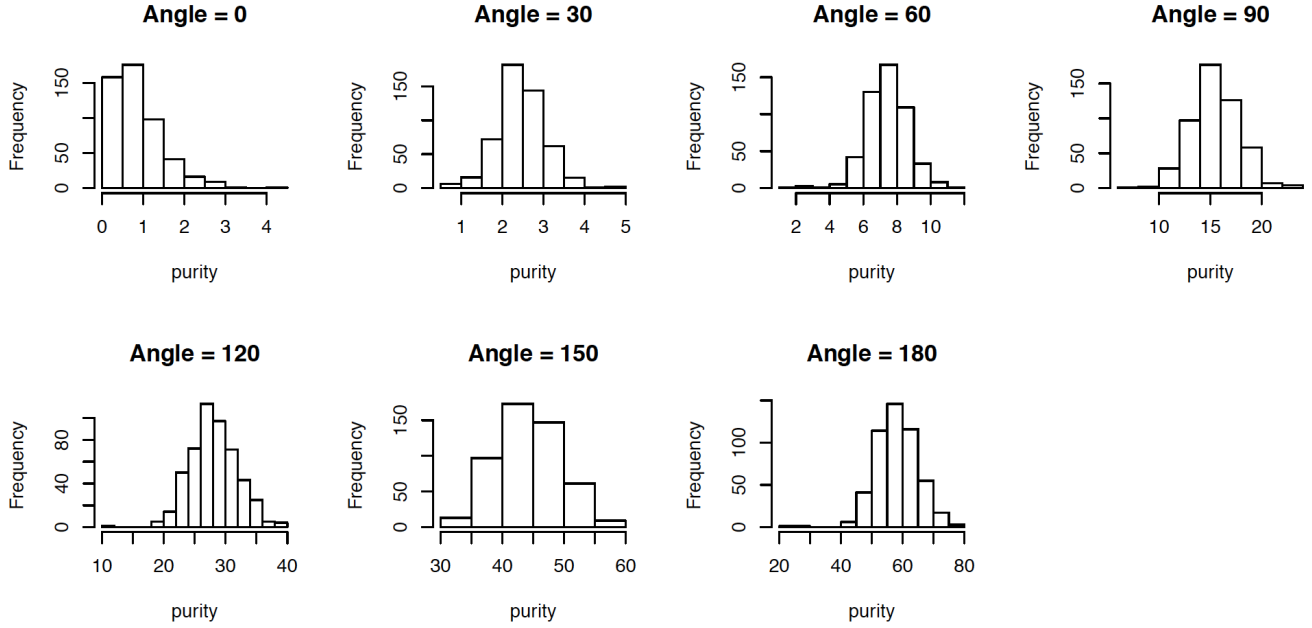
$\alpha$  calculated by  $\hat{\beta}_{drg}, \hat{\beta}_{pbo}, \hat{\Gamma}_{drg}, \hat{\Gamma}_{pbo}$ , true  $D_{drg}, D_{pbo}$

Angle	Mean KL dif	Mean KL	SD KL	Mean cosine	SD cosine
0	0.88	0.88	0.60	0.28	0.41
30	0.46	2.46	0.58	0.62	0.67
60	0.82	7.40	1.24	0.67	0.72
90	1.50	15.45	2.28	0.66	0.75
120	2.38	28.04	3.79	0.67	0.74
150	3.16	44.26	5.25	0.67	0.74

Angle	Mean KL dif	Mean KL	SD KL	Mean cosine	SD cosine
180	3.75	58.16	6.71	0.67	0.74

- Mean\_KL\_dif: the mean absolute differences between true *purity* and the estimated  $\hat{purity}$  during the 500 times replication.
- Mean\_KL: The mean value of the estimated *purity*
- SD\_KL: The standard deviation value of the estimated *purity*
- Mean\_cosine: the mean value of the cosine\_similarity.
- SD\_cosine: The standard deviation value of the cosine\_similarity

### The histogram of purity



### The histogram of the consine similarity

