Simulation to estimating alpha 3

2019-09-15

Contents

Kullback-Leibler divergence

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

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

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

Therefore, we defined the *puirty* as

$$\int log(f_1)f_1 - \int log(f_2)f_1 + \int log(f_2)f_2 - \int log(f_1)f_2$$

$$= \Big(-\frac{p}{2}log(2\pi) - \frac{1}{2}log(|D_1|) - \frac{p}{2} \Big)$$

$$- \Big(-\frac{p}{2}log(2\pi) - \frac{1}{2}log(|D_2|) - \frac{1}{2}\Big(tr(D_2^{-1}D_1) + (\mu_1 - \mu_2)'D_2^{-1}(\mu_1 - \mu_2) \Big) \Big)$$

$$+ \Big(-\frac{p}{2}log(2\pi) - \frac{1}{2}log(|D_2|) - \frac{p}{2} \Big)$$

$$- \Big(-\frac{p}{2}log(2\pi) - \frac{1}{2}log(|D_1|) - \frac{1}{2}\Big(tr(D_1^{-1}D_2) + (\mu_1 - \mu_2)'D_1^{-1}(\mu_1 - \mu_2) \Big) \Big)$$

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

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

That is:

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

The optimization of α:

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

where

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

- $B = \sum_i x_i\Big((\Gamma_1 - \Gamma_2)\Big)'(D_1^{-1} + D_2^{-1})\Big((\Gamma_1 - \Gamma_2)\Big)x_i'$

Simulation Setting

Old results (200 subjects each arm)

Sample size:

- 200 subjects in each group, 400 in total.

Dimensions:

- p = 4. (4 baseline biosignature)

Angle between Γ_1 and Γ_3:

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

Initial value: True α

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 Γ lines can be 0, 30, 60, 90, 120, 150, and 180 degrees.

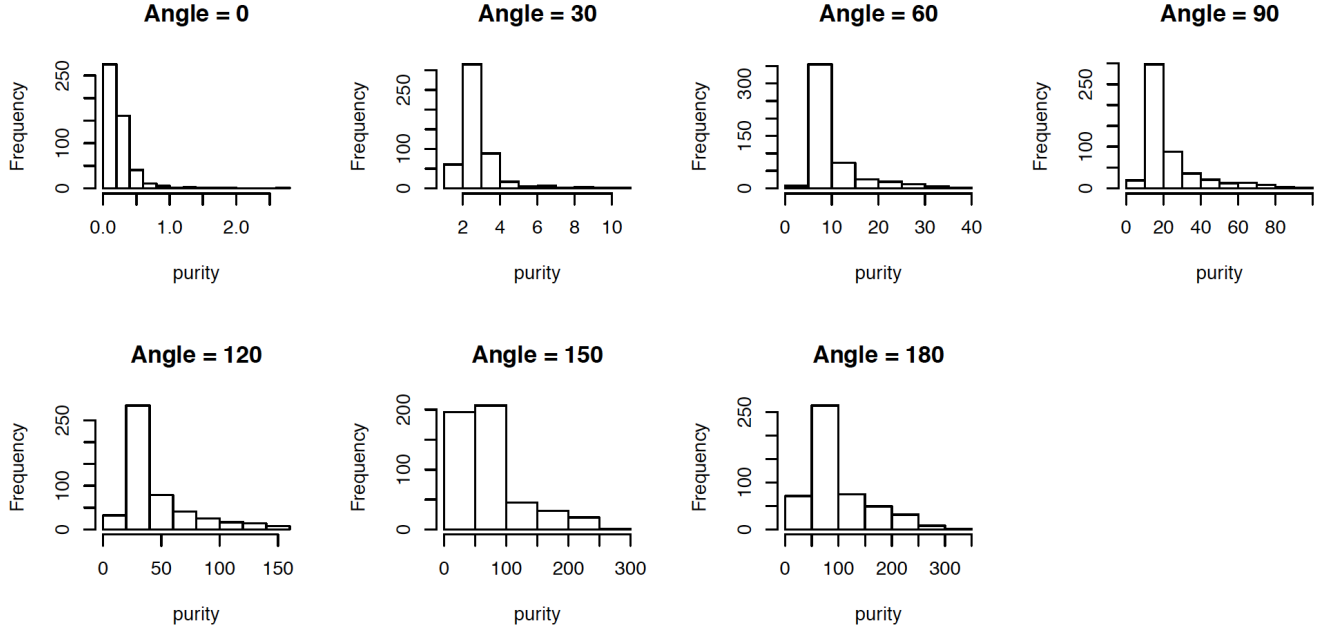- $|\Gamma_{drg}| = |\Gamma_{pbo}| = 1$

α 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	Mean abs cosine	SD abs cosine
0	0.24	0.24	0.24	0.35	0.54	0.55	0.33
30	0.70	2.79	1.07	0.66	0.73	0.98	0.02
60	3.37	10.18	5.64	0.66	0.72	0.98	0.05
90	9.47	22.76	15.26	0.65	0.73	0.97	0.09
120	21.22	44.90	30.80	0.64	0.72	0.96	0.10
150	35.49	75.20	49.32	0.67	0.73	0.99	0.02
180	44.41	96.89	57.38	0.66	0.74	0.99	0.02

- Mean_KL_dif: the mean absolute differences between true *purity* and the estimated *purîty* 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

- Mean_abs_cosine: the mean value of the absolute cosine_similarity.

- SD_abs_cosine: The standard deviation value of the absolute cosine_similarity
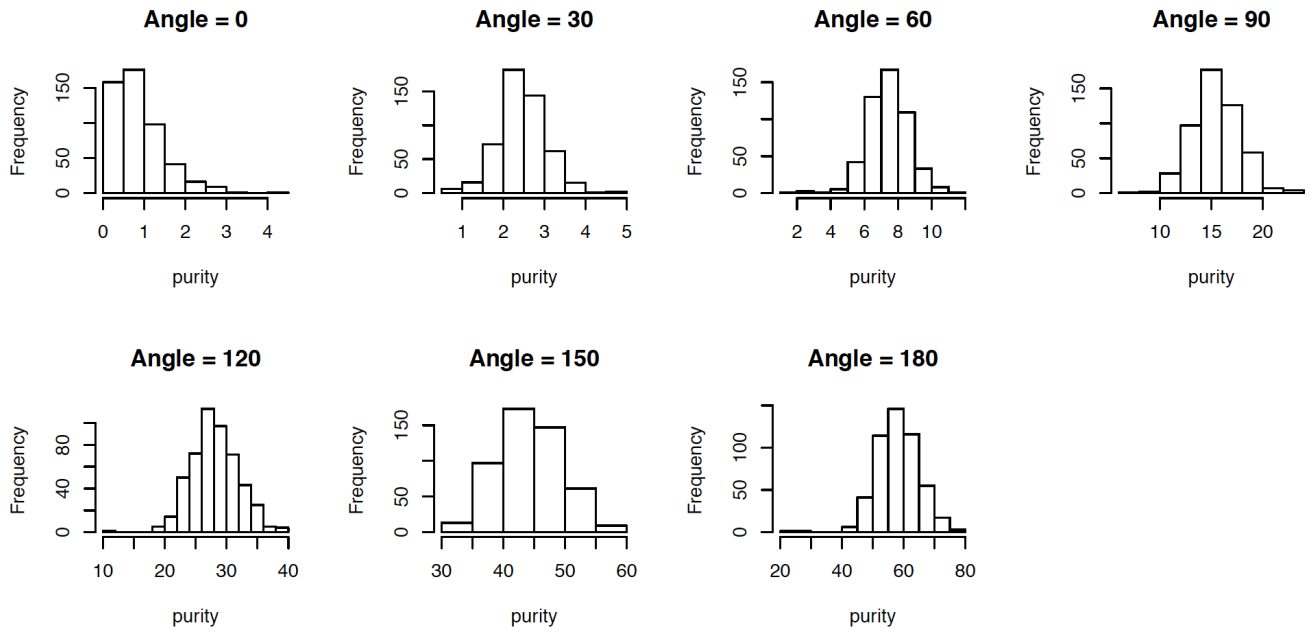
The histogram of purity

α **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	Mean abs cosine	SD abs cosine
0	0.88	0.88	0.60	0.28	0.41	0.43	0.27
30	0.46	2.46	0.58	0.62	0.67	0.89	0.21
60	0.82	7.40	1.24	0.67	0.72	0.99	0.07
90	1.50	15.45	2.28	0.66	0.75	0.99	0.02
120	2.38	28.04	3.79	0.67	0.74	0.99	0.02
150	3.16	44.26	5.25	0.67	0.74	1.00	0.01
180	3.75	58.16	6.71	0.67	0.74	1.00	0.02

- 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

- Mean_abs_cosine: the mean value of the absolute cosine_similarity.

- SD_abs_cosine: The standard deviation value of the absolute cosine_similarity

The histogram of purity
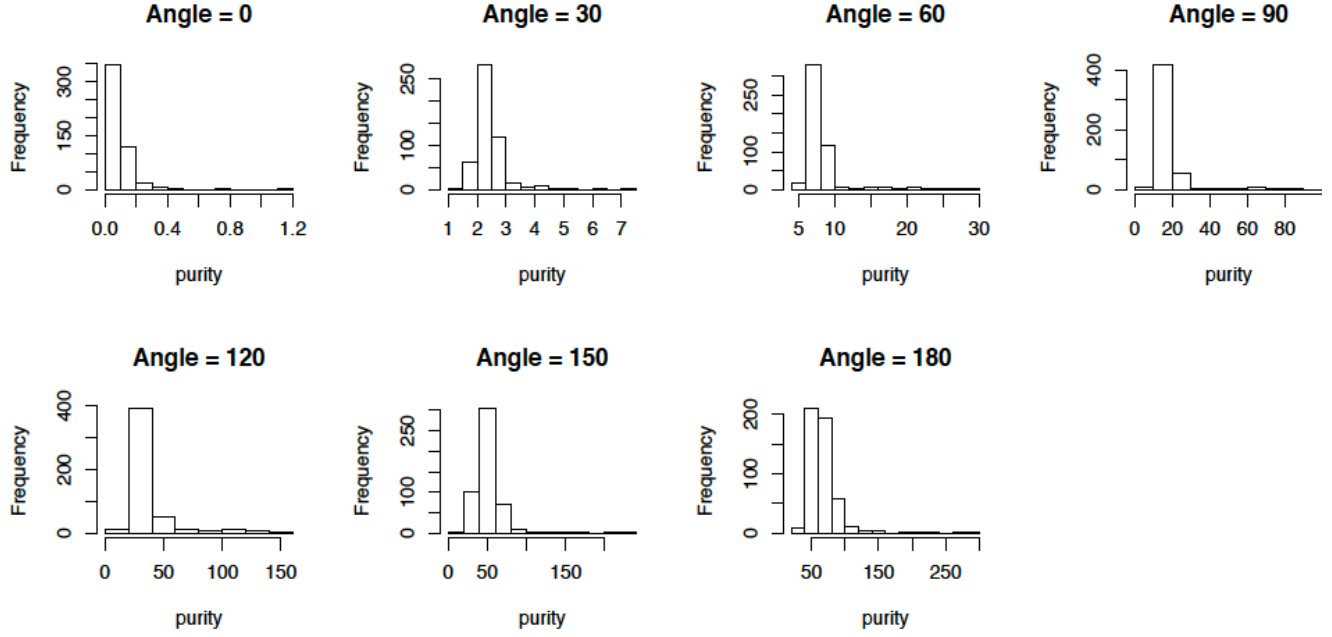
New results (500 each arm)

Sample size:

- 500 subjects in each group, 1000 in total.

α 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	Mean abs cosine	SD abs cosine
0	0.09	0.10	0.10	0.35	0.53	0.55	0.32
30	0.33	2.42	0.55	0.73	0.67	0.99	0.01
60	1.44	8.22	3.22	0.74	0.66	0.99	0.03
90	4.09	17.84	10.46	0.73	0.67	0.99	0.03
120	11.07	36.31	21.97	0.72	0.67	0.98	0.06
150	11.51	51.44	21.27	0.73	0.68	1.00	0.01
180	16.24	68.93	30.72	0.72	0.69	1.00	0.01

- Mean_KL_dif: the mean absolute differences between true *purity* and the estimated *purîty* 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

- Mean_abs_cosine: the mean value of the absolute cosine_similarity.

- SD_abs_cosine: The standard deviation value of the absolute cosine_similarity

4

The histogram of purity

α 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	Mean abs cosine	SD abs cosine
0	0.27	0.27	0.22	0.61	0.31	0.62	0.29
30	0.26	2.28	0.31	0.99	0.01	0.99	0.01
60	0.58	7.23	0.69	1.00	0.00	1.00	0.00
90	1.12	15.11	1.32	1.00	0.00	1.00	0.00
120	1.88	27.60	2.17	1.00	0.00	1.00	0.00
150	2.73	43.69	3.09	1.00	0.00	1.00	0.00
180	3.40	57.61	3.79	1.00	0.00	1.00	0.00

- 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

- Mean_abs_cosine: the mean value of the absolute cosine_similarity.

- SD_abs_cosine: The standard deviation value of the absolute cosine_similarity

The histogram of purity