Simulation set up

A single optimal cut-off point was identified using the Youden method within a Bayesian classification framework, applied to a bivariate mixture model of CSF and plasma. The amyloid positive was define as z $>$ 0.5. Subject was label as amyloid positive when k = 2.

$$f\begin{pmatrix} y^{\text{csf}} \\ y^{\text{plasma}} \end{pmatrix} = \sum_{k=1}^{2} \pi_k \cdot N\left(\ \mu_k\ , \Sigma\right)$$

$$\text{where } \mu_1 = (0.05, 0.08), \mu_2 = (0.1, 0.1), \Sigma = \begin{pmatrix} 0.0001082 & 0.0000375 \\ 0.0000375 & 0.0001030 \end{pmatrix}, \pi_1 = 0.3, \pi_2 = 0.7$$

Findings

A single optimal cut-off point was identified using the Youden method within a Bayesian framework classification, applied to a bivariate mixture model of CSF and plasma. Youden method select the cut-point maximizing Youden's J statistic which is sensitivity + specificity -1. For each sample size, the average cut-off point, sensitivity, specificity, accuracy were calculated across 1,000 datasets. However, I have yet to resolve why the sensitivity, specificity, accuracy, and AUC values decrease as the sample size increases.

	optimal_cutpoint	Accuracy	sensitivity	specificity
real data with sample size 130	0.0915	0.8231	0.8182	0.8387
simulated data with sample size 100	0.0904 (0.0026)	0.8550 (0.0379)	0.8400 (0.0631)	0.8900 (0.0847)
simulated data with sample size 200	0.0898 (0.0023)	0.8523 (0.0319)	0.8488 (0.0553)	0.8607 (0.0579)
simulated data with sample size 500	0.0899 (0.0017)	0.8450 (0.0217)	0.8442 (0.0415)	0.8470 (0.0450)

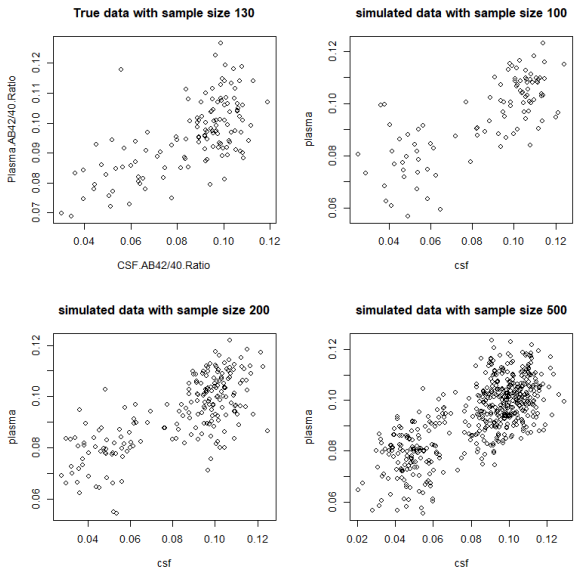

Figure 1: scatter plot of real data and one sample simuation data of each sample size

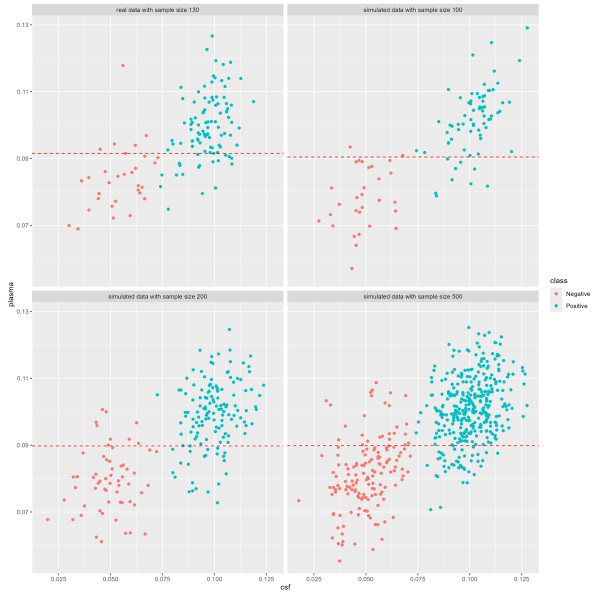

Figure 2: scatter plot of real data and one sample simuation data of each sample size with classification