

3 Part Three - Simulations

Nella cartella Drive trovate il file *nuove_funzioni.R*

Structural Learning:

- Iniziate tenendo la partizione fissata
- Per ottenere la stima a posteriori della matrice di precisione, usate la formula 15 del mio paper (Codazzi, Colombi Gianella, 2021).
- Per valutare la bontà della stima della matrice di precisione, usate la Kullback Liebler divergence. Trovate la funzione nel file R.
- Per il grafo, trovate la posterior probability of inclusion matrix (*plinks*, trovate la funzione in BDgraph. Se non riuscite ad usare quella funzione perché ha bisogno di un input particolare, guardate la formula 16 sempre dello stesso paper)
- Come grafo finale, avete due opzioni:
 - Include tutti quei lati per cui $\hat{p}_{ij} \geq 0.5$. Spesso questo modo funziona male.
 - Usate il criterio Bayesian False Discovery Rate (BFDR), trovate la funzione del file R

Clustering:

- Parametizzate la Beta con media pari alla densità vera della rete e varianza 1/16
- Installate i pacchetti `salso`, `mcclust` e `mcclust.ext` (<https://github.com/sarawade/mcclust.ext>). Vedere file su quali funzioni usare e come
- Ad ogni iterazione, calcolare il rand index tra la partizione corrente e quella vera. Fate un traceplot alla fine
- Usate la Binder loss e la Variation of Information per ottenere una sintesi finale della partizione. Controllate che la partizione finale rispetti l'ordinamento desiderato
- Calcolate rand index della partizione finale
- Calcolate il numero di cluster stimati sia come stima monte carlo (media del numero di cluster iterazione per iterazioni) che sia come il numero di cluster della partizione finale. Di solito, quello calcolato iterazione per iterazione è peggiore
- Fate istogramma e traceplot del numero di cluster ottenuti durante il sampling
- Fate un istogramma della probabilità che ogni nodo sia o no un changepoint. Ovvero, ad ogni iterazione avete il vettore z in cui gli elementi sono 1 se il nodo è CP e 0 altrimenti. Fate l'istogramma di quelle z . Per intenderci, voglio replicare il panel basso della Fig. 2 del paper Corradin-Danese
- Quanto cambiano i risultati al variare dei parametri della Beta?
- Ripetete un certo numero di volte la simulazione cambiando ogni volta i dati. Generateli sempre nello stesso modo ma cambiando il seed.
 - Salvate Rand Index e numero di cluster stimati per ogni dataset. Per entrambi gli indici, fate un boxplot e calcolate media/mediana/standard deviation

References