0.1 Variational decomposition and mean field approximation

The likelihood of an observation under the SBM, in Equation (**??**), is complex to deal with for mainly two reasons. First, it is a sum over all latent configurations, and thus it has an exponential number of terms, making it intractable. Second, it can have multiple local optima. Therefore, approximations are needed in order to work with this model. A common one is the variational approximation to the likelihood, which will deal with the problem of summing an exponential number of terms. This is still complicated in all its generality, so a second "mean field" approximation is used on top of the first variational one. This consists in searching the solution to the variational approximation amidst factorizable distributions.

0.1.1 Deriving the variational decomposition.

Let Z denote the vector of latent variables, A an observation, and θ the parameters of an SBM. The following *variational decomposition* leads to an useful approximation to the likelihood.

Definition 1. For any two distributions $p(z), q(z)$,

$$\mathrm{KL}(q\|p) := -\int_Z \log\left(\frac{p(z)}{q(z)}\right) q(z)\, dz \tag{1}$$

is called the *Kullback-Leibler divergence* from $p(z)$ to $q(z)$.

Theorem 1. *The observed likelihood can be decomposed as*

$$\log p(A;\theta) = \mathcal{L}(q,\theta) + \mathrm{KL}(q\|p(Z|A;\theta)), \tag{2}$$

where

$$\mathcal{L}(q,\theta) := \int_Z \log\left(\frac{p\,(Z,A;\theta)}{q(Z)}\right) q(Z)\, dZ \tag{3}$$

is called the evidence lower bound, or ELBO for short.

Proof. The posterior is defined as

$$p\,(Z|A;\theta) = \frac{p\,(Z,A;\theta)}{p\,(A;\theta)},$$

which implies

$$p\,(A;\theta) = \frac{p\,(Z,A;\theta)}{p\,(Z|A;\theta)}.$$

Thus,

$$\log p(A;\theta) = \log p(Z, A; \theta) - \log p(Z|A;\theta).$$

Taking the expectation of this expression with respect to some proposal distribution $q(Z)$ depending only on Z, one has

$$\begin{aligned}
\log p(A;\theta) &= \int_Z \log p(Z, A;\theta)\, q(Z)\ dZ - \int_Z \log p(Z|A;\theta)\, q(Z)\ dZ \\
&= \int_Z \left(\log\left(\frac{p(Z, A;\theta)}{q(Z)}\right) - \log q(Z) \right) q(Z)\ dZ \\
&\quad - \int_Z \left(\log\left(\frac{p(Z|A;\theta)}{q(Z)}\right) - \log q(Z) \right) q(Z)\ dZ.
\end{aligned}$$

Therefore,

$$\log p(A;\theta) = \int_Z \log\left(\frac{p(Z, A;\theta)}{q(Z)}\right) q(Z)\, dZ - \int_Z \log\left(\frac{p(Z|A;\theta)}{q(Z)}\right) q(z)\, dZ,$$

that is,

$$\log p(A;\theta) = \mathcal{L}(q,\theta) + \mathrm{KL}(q \| p(Z|A;\theta)).$$

□

Equation (2) forms the basis of the classical EM algorithm for estimating θ, where one performs alternate minimization of the KL term and subsequent maximization of the term $\mathcal{L}(q,\theta)$. A classical result proves that the KL is always positive. Therefore, the ELBO is indeed a lower bound for the log likelihood being decomposed. Given the independence of the left hand side with respect to q, observe that maximizing the ELBO amounts to minimizing the KL term, and this can be done by setting $q = p(Z|A;\theta)$. However, in the case of the SBM, this conditional probability is itself intractable, therefore this step of EM must be performed differently.

The ELBO is also called the "free energy" in the literature.

0.1.2 Mean field approximation.

A common strategy used to deal with the problem of having an untractable solution q to the variational approximation is called the "mean field approximation". It consists in trying to find a q distribution maximizing the ELBO constrained to a family of tractable distributions. Here, tractability means factorizability: consider distributions of the form

There are other "correlated" mean field approximations, see [?].

$$q(Z) = \prod_{i=1}^{n} q_i(Z_i).$$

In the case of the SBM, each factor must be multinomial distribution, and they differ only by their parameters, that is,

$$q(Z) = \prod_{i=1}^{n} m(Z_i; \tau_i), \tag{4}$$

where $m(\cdot, \tau_i)$ is the probability mass function of a multinomial distribution with parameter τ_i.

Proposition 1. *The mean-field ELBO is given by*

$$\begin{aligned}\mathcal{L} = \sum_{i=1}^{n}\sum_{k=1}^{K}\Bigg[&\tau_{ik}\log\frac{\pi_k}{\tau_{ik}} \\ &+\frac{1}{2}\sum_{j=1}^{n}\sum_{l=1}^{K}\tau_{ik}\tau_{jl}\left(A_{ij}\log\gamma_{kl}+(1-\delta_{ij}-A_{ij})\log(1-\gamma_{kl})\right)\Bigg].\end{aligned} \tag{5}$$

Proof. Similarly to [?], substitute Equation (4) into the ELBO of Equation (2), to get

$$\begin{aligned}\mathcal{L} = \sum_{i=1}^{n}\sum_{k=1}^{K}\mathbb{E}_q[Z_{ik}]\log\pi_k &+ \sum_{j>i}A_{ij}\mathbb{E}_q\left[\log\gamma_{Z_i,Z_j}\right] \\ &+ (1-A_{ij})\,\mathbb{E}_q\left[\log\left(1-\gamma_{Z_i,Z_j}\right)\right]+\mathcal{H}(q),\end{aligned} \tag{6}$$

where $\mathcal{H}(q) = -\sum_Z q(Z)\log(Z)$ is the entropy of the distribution q. This expression can be further simplified by noticing that the expectations appearing in it can be explicitly calculated. First, notice that $\mathbb{E}_q[Z_{ik}] = \tau_{ik}$. Also notice that for each i, j,

$$\begin{aligned}\mathbb{E}_q\left[\log\gamma_{Z_i,Z_j}\right] &= \sum_Z q(Z)\log\gamma_{Z_i,Z_j} \\ &= \sum_{Z_i,Z_j} q(Z_i,Z_j)\log\gamma_{Z_i,Z_j} \\ &= \sum_{Z_i,Z_j} m(Z_i,\tau_i)\,m(Z_j,\tau_j)\log\gamma_{Z_i,Z_j} \\ &= \sum_{Z_i,Z_j}\prod_{k=1}^{K}\tau_{ik}^{Z_{ik}}\prod_{l=1}^{K}\tau_{jl}^{Z_{jl}}\log\gamma_{Z_i,Z_j} \\ &= \sum_{k=1}^{K}\sum_{l=1}^{K}\tau_{ik}\tau_{jl}\log\gamma_{kl}.\end{aligned}$$

Similarly, $\mathbb{E}_q[\log(1-\gamma_{Z_i,Z_j})] = \sum_{k,l}\tau_{ik}\tau_{jl}\log(1-\gamma_{kl})$. The entropy term can also be simplified by noticing that

$$\begin{aligned}
\mathcal{H}(q) &= \mathcal{H}\left(\prod_{i=1}^{n} m(Z_i;\tau_i)\right) \\
&= -\sum_{Z}\left(\prod_{i=1}^{n} m(Z_i;\tau_i)\right)\log\prod_{j} m(Z_j;\tau_j) \\
&= -\sum_{j=1}^{n}\sum_{Z}\left(\prod_{i=1}^{n} m(Z_i;\tau_i)\right)\log m(Z_j;\tau_j) \\
&= -\sum_{j=1}^{n}\sum_{Z_j} m(Z_j,\tau_j)\log m(Z_j,\tau_j) \\
&= \sum_{j=1}^{n}\mathcal{H}(m(Z_j,\tau_j)),
\end{aligned}$$

where the sum in Z became a sum in Z_j by marginalizing out the remaining multinomial mass functions. Each such term can be calculated:

$$\begin{aligned}
\mathcal{H}(m(Z_j,\tau_j)) &= -\sum_{Z_j} m(Z_j,\tau_j)\log m(Z_j,\tau_j) \\
&= -\sum_{Z_j}\left(\prod_{i=1}^{K}\tau_{ji}^{Z_{ji}}\right)\sum_{k=1}^{K} Z_{jk}\log\tau_{jk} \\
&= -\sum_{i=1}^{K}\tau_{ji}\log\tau_{ji},
\end{aligned}$$

where the expression is simplified by considering that the possible values for Z_j are the different indices where its only non-zero component can be. Substituting all of this back into Equation (6), one finds that the expression for the mean-field ELBO, as in the theorem statement. $\square$

0.1.3 Variational estimators

Given edge observations A_{ij}, the mean-field parameters $\hat{\tau} = (\hat{\tau}_{ik})_{\substack{i=1,\dots,n\\k=1,\dots,K}}$ and the SBM parameters $\hat{\theta} = (\hat{\pi}_k, \hat{\gamma}_{kl})_{k,l=1,\dots,K}$ can be estimated by maximizing 5 in alternate steps. Having fixed a value for $\hat{\theta}$, maximize the mean-field ELBO with respect to τ under the constraint that $\tau_{ik} \geq 0$ for $1 \leq i \leq n$ and $1 \leq k \leq K$, and that $\sum_k \tau_{ik} = 1$ for

$1 \leq i \leq n$; then, for a fixed $\hat{\tau}$, maximize the mean-field ELBO with respect to θ under the constraint $0 < \pi_k < 1, 1 \leq k \leq K$ and so on.

Optimizing for $\hat{\tau}$

By penalizing the constraints for τ on the ELBO, one gets a Lagrangian $\mathscr{L}$ which when derived and equaled to zero yields

$$\begin{aligned}\nabla_{\tau_{ik}} \mathscr{L} = \log \pi_k - 1 - \log \tau_{ik} + \sum_{\substack{j>i \\ l=1,\ldots,K}} & A_{ij} \tau_{jl} \log \gamma_{kl} \\ & + (1 - A_{ij})\, \tau_{jl} \log (1 - \gamma_{kl}) + \mu_i \\ = 0. &\end{aligned}$$

This can be directly rearranged to the following fixed point relation

$$\hat{\tau}_{ik} \propto \pi_k \prod_{\substack{j>i \\ l=1,\ldots,K}} \left(\gamma_{kl}^{A_{ij}} \left(1 - \gamma_{kl}\right)^{(1-A_{ij})} \right)^{\tau_{jl}}. \tag{7}$$

Thus, one "general way" of obtaining $\hat{\tau}$ is by evaluating Equation (7) repeatedly until convergence.

Optimizing for $\hat{\theta}$

As π_k appears inside a logarithm in the ELBO, the positivity constraint is naturally enforced by the objective function. Therefore one only needs to impose that it sums to one. The Lagrangian then becomes

$$\mathscr{L} = -\sum_{i=1}^{n} \sum_{k=1}^{K} \tau_{ik} \log \frac{\pi_k}{\tau_{ik}} + \mu \left(\sum_{l=1}^{K} \pi_l - 1 \right),$$

and its derivative equaled to zero provides the equation

$$\nabla_{\pi_k} \mathscr{L} = -\sum_{i=1}^{n} \frac{\tau_{ik}}{\pi_k} + \mu = 0.$$

This optimality condition gives the estimator sought in terms of the Lagrange multiplier

$$\hat{\pi}_k = \frac{1}{\mu} \sum_{i=1}^{n} \tau_{ik}.$$

To find the value of the Lagrange multiplier, one must solve the dual problem. Lagrange's dual function writes

$$q(\mu) = -\sum_{i=1}^{n}\sum_{k=1}\tau_{ik}\log\left(\frac{\sum_{l=1}^{K}\tau_{lk}}{\mu\tau_{ik}}\right) + \tau_{ik} - \mu.$$

The dual problem is $\max_\mu q(\mu)$, which is unconstrained:

$$\nabla_\mu q = \sum_{i=1}^{n}\sum_{k=1}^{K}\frac{\tau_{ik}}{\mu} - 1 = 0 \implies \mu = \sum_{i=1}^{n}\sum_{k=1}^{K}\tau_{ik} = n.$$

Substituting the value found for μ back into the estimator for $\hat{\pi}$, one concludes

$$\hat{\pi}_k = \frac{1}{n}\sum_{i=1}^{n}\tau_{ik}. \tag{8}$$

A similar procedure generates estimators for the connectivities γ, the difference being that these are unconstrained, so that by deriving the ELBO directly one gets

$$\begin{aligned}\nabla_{\gamma_{kl}}\mathcal{L} &= \sum_{i=1}^{n}\sum_{j=1}^{n}\tau_{ik}\tau_{jl}\left(\frac{A_{ij}}{\gamma_{kl}} - \frac{1-\delta_{ij}-A_{ij}}{1-\gamma_{kl}}\right) = 0\\ &\implies \sum_{i=1}^{n}\sum_{j=1}^{n}\tau_{ik}\tau_{jl}\left(A_{ij} - \gamma_{kl}\left(1-\delta_{ij}\right)\right) = 0.\end{aligned}$$

This optimality condition yields the estimator sought:

$$\hat{\gamma}_{kl} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\tau_{ik}\tau_{jl}A_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{n}\tau_{ik}\tau_{jl}\left(1-\delta_{ij}\right)}. \tag{9}$$

These equations can be simplified in the simple case of two communities, which is a fundamental example for building intuition and testing hypotheses.

0.1.4 Convergence properties