Flexible sampling of a pair of correlated positive integers

Richard Yi Da Xu

University of Technology Sydney (UTS)

July 9, 2017

徐亦达

wechat ID: **aubedata**

Notes: `http://www-staff.it.uts.edu.au/~ydxu/statistics.htm`

So what is Copula

- A bivariate copula function $C(u, v)$ is a Cumulative Distribution Function over the interval $[0, 1] \times [0, 1]$ with uniform marginal distribution.
- **Sklar's theorem:**
  Let X and Y be random variables with **marginal** distribution F and G and **joint** distribution $\Pr(.)$. Then there exists a Copula C such that for all $(x, y) \in R \times R$:

$$\Pr(x, y) = C(F(x), G(y))$$

- C is unique if F and G are continuous, then the joint probability density function is:

$$\begin{aligned} p(x, y) =& \frac{\partial C(F(x), G(y))}{\partial x \partial y} \\ =& c(F(x), G(y)) \frac{\partial F(x)}{\partial x} \frac{\partial G(y)}{\partial y} \\ =& c(F(x), G(y)) f(x) g(y) \end{aligned}$$

- Here $c(u, v) = \frac{\partial^2 C(u,v)}{\partial u \partial v}$ is copula density function.

Some goodies about Copula

- Sklar's theorem ensures the uniqueness of copula function $C(F(x), G(y))$
- Change Copula function does not change the marginal distributions. This can be a very interesting property.
- Copula is popular! Many are availability to **suit the situation**. Commonly used copula functions includes, Gaussian Copula (Gaussian, t), Archimedean Copula (Clayton, Gumbel, Frank, etc.), Empirical Copula.

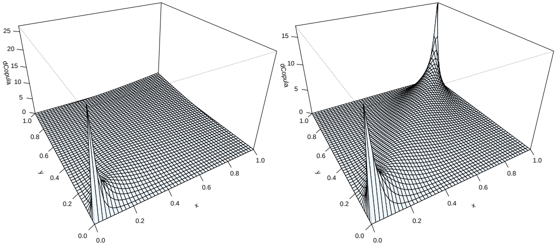


Figure: Clayton Copula ($\theta = 2$) and Gaussian Copula ($\theta = 0.9$) visualization

Generative Model

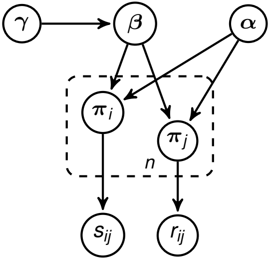
- C1: $\boldsymbol{\beta} \sim GEM(\gamma)$
- C2: $\pi_i, \pi_j \sim DP(\alpha \cdot \boldsymbol{\beta})$
- C4: $s_{ij} \sim \text{Mult}(\pi_i) \qquad r_{ji} \sim \text{Mult}(\pi_j)$
  alternatively:

$$s_{ij} = \Pi_i^{-1}(u) \qquad r_{ij} = \Pi_j^{-1}(v)$$
$$u \sim U(0,1) \qquad v \sim U(0,1)$$

$$\Pi_i^{-1}(u) = \{\min q : \sum_{q=1}^{k} \pi_{iq} \geq u\}$$

$$\Pi_j^{-1}(v) = \{\min q : \sum_{q=1}^{k} \pi_{jq} \geq v\}$$

Graphical model



Copula based DP samples

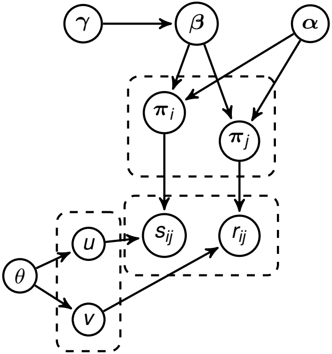
- In HDP, s_{ij} and r_{ji} are sampled independently.
- Our intention is to design a way to model correlations between s_{ij} and r_{ji}.
- However, we need to leave marginal distributions of s_{ij} and r_{ji} invariant.
- You guessed it: use Copula!

Generative Model

- C1: $\boldsymbol{\beta} \sim GEM(\gamma)$
- C2: $\pi_i, \pi_j \sim DP(\alpha \cdot \boldsymbol{\beta})$
- C4: $s_{ij} \sim \text{Mult}(\pi_i) \qquad r_{ji} \sim \text{Mult}(\pi_j)$
  alternatively:

$$s_{ij} = \Pi_i^{-1}(u) \qquad r_{ij} = \Pi_j^{-1}(v)$$
$$u \sim U(0,1) \qquad v \sim U(0,1)$$
$$\Pi_i^{-1}(u) = \{\min q : \sum_{q=1}^{k} \pi_{iq} \geq u\}$$
$$\Pi_j^{-1}(v) = \{\min q : \sum_{q=1}^{k} \pi_{jq} \geq v\}$$

Graphical model

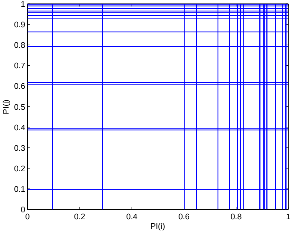


Diagrammatic Representation

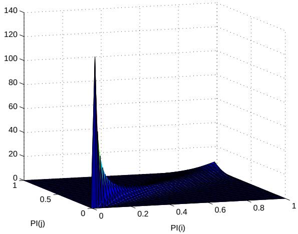
In words, it says:

- Get two independent Dirichlet Processes draws, π_i and π_j to form the 2D grid.
- Get 2D random sample $\in [0 \dots 1]^2$ using $(u, v)^\top \sim$ Copula(θ) function
- Then to compute **deterministically** which grid this 2D sample falls into.

- $\pi_i, \pi_j \sim DP(\alpha \cdot \boldsymbol{\beta})$



- $(u, v) \sim$ Copula(θ)



Why Bivariate Copula-DP a generalisation of HDP?

All we need is to substitute the copula with uniform distribution:

$$u_{ij} \sim U(0,1), v_{ij} \sim U(0,1) \qquad s_{ij} = \Pi_i^{-1}(u_{ij}), r_{ij} = \Pi_j^{-1}(v_{ij}).$$

Inference of Copula-MMSB

- Standard Gibbs Sampling **does the job**, but mixes slowly.
- We desire a collapsed Gibbs sampling for faster mixing on Markov chains.
- An "Ideal" analytical expression for marginal distribution of $\Pr(s_{ij}, r_{ij})$ can NOT be achieved:

$$\Pr(s_{ij}, r_{ij}) = \int_{\sum_{d=1}^{K+1} \pi_{jd}=1} \int_{\sum_{d=1}^{K+1} \pi_{id}=1} \int_{(u,v)} \cdot \mathbf{1}\left(s_{ij} = \Pi_i^{-1}(u), r_{ij} = \Pi_j^{-1}(v)\right) \cdot \mathrm{d}C(u,v)\mathrm{d}F(\pi_{i1}, \cdots, \pi_{iK+1})\mathrm{d}F(\pi_{j1}, \cdots, \pi_{jK+1})$$

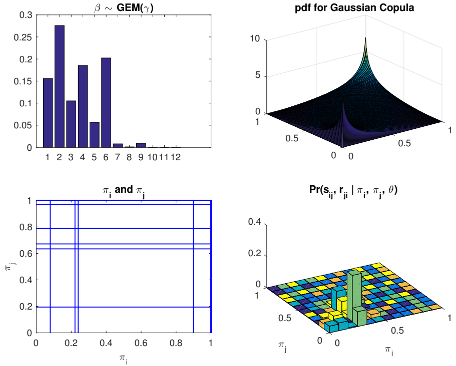
- With some mathematical design, conditioning on samples of either (u, v) or (π_i, π_j), it is possible to obtain a marginalised conditional density in which s_{ij}, r_{ij} is conditioned on either (u, v) or (π_i, π_j), but not both.
- Therefore, two inference schemes are presented:
  - **Marginal conditional on π only**
  - **Marginal conditional on u, v only**

Marginal Conditional on π only

- Let $C(u, v|\theta)$ be the chosen Copula cumulative distribution function (c.d.f.) with parameter θ

$$\Pr(s_{ij}, r_{ij}|\pi_i, \pi_j, \theta) = \int_{\hat{\pi}_i^{k-1}}^{\hat{\pi}_i^k} \int_{\hat{\pi}_j^{l-1}}^{\hat{\pi}_j^l} \mathrm{d}C(u, v|\theta) \qquad \hat{\pi}_i^k = \begin{cases} 0, & k = 0; \\ \sum_{q=1}^k \pi_{iq}, & k > 0 \end{cases}$$

$$= C(\hat{\pi}_i^k, \hat{\pi}_j^l) + C(\hat{\pi}_i^{k-1}, \hat{\pi}_j^{l-1}) - C(\hat{\pi}_i^k, \hat{\pi}_j^{l-1}) - C(\hat{\pi}_i^{k-1}, \hat{\pi}_j^l)$$



Marginal Conditional on u and v only

- Given u and v, $\Pr(s_{ij})$ and $\Pr(s_{ij})$ are independent.

$$\Pr(s_{ij} = k, r_{ij} = l | \_\_) \propto \Pr(s_{ij} = k | u, \alpha, \boldsymbol{\beta}) \cdot \Pr(r_{ij} = l | v, \alpha, \boldsymbol{\beta})$$

- Each in the form of $\Pr(s_{ij} = k | u_{ij}, \lambda_1, \dots, \lambda_K)$
- (later) - when having a multinomial likelihood, conjugacy ensures it is still in this form.
- Using properties:

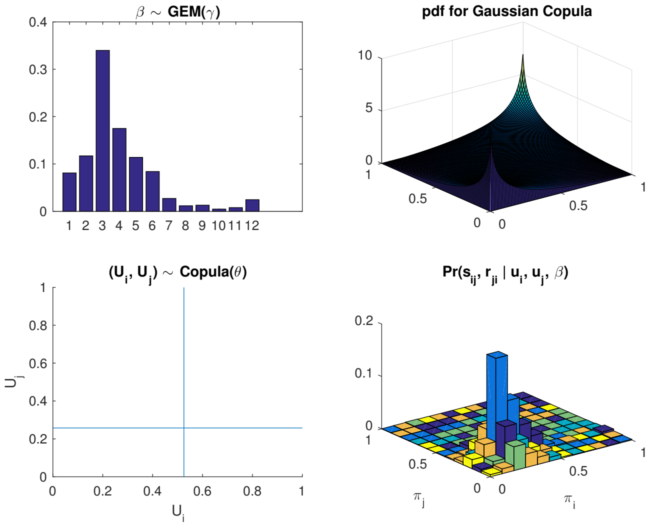
$$\pi_{i1}, \dots, \pi_{i,K} \sim \text{DIR}(\lambda_1, \dots, \lambda_K)$$
$$\implies \Pr\left(\sum_{d=1}^{k} \pi_{id} < u\right) = \text{Betacdf}\left(\sum_{d=1}^{k} \lambda_d, \sum_{d=k+1}^{K} \lambda_d\right) \triangleq \text{Betacdf}(h^k, \hat{h}^k)$$

- Therefore, $\forall d$, $d \neq 1$ and $d \neq k$:

$$\Pr\left(\sum_{d=1}^{k-1} \pi_{id} \leq u_{ij} \bigcap \sum_{d=1}^{k} \pi_{id} > u_{ij}\right) = 1 - \Pr\left(\sum_{d=1}^{k-1} \pi_{id} \geq u_{ij} \bigcup \sum_{d=1}^{k} \pi_{id} < u_{ij}\right)$$
$$= 1 - \Pr\left(\sum_{d=1}^{k-1} \pi_{id} \geq u_{ij}\right) - \Pr\left(\sum_{d=1}^{k} \pi_{id} < u_{ij}\right) + \Pr\left(\sum_{d=1}^{k-1} \pi_{id} \geq u_{ij} \bigcap \sum_{d=1}^{k} \pi_{id} < u_{ij}\right)$$
$$= 1 - [1 - \text{Betacdf}_{u_{ij}}(h_i^{k-1}, \hat{h}_i^{k-1})] - \text{Betacdf}_{u_{ij}}(h_i^k, \hat{h}_i^k) + \emptyset$$
$$= \text{Betacdf}_{u_{ij}}(h_i^{k-1}, \hat{h}_i^{k-1}) - \text{Betacdf}_{u_{ij}}(h_i^k, \hat{h}_i^k)$$

- In this context $\lambda_d = \alpha\beta_d$

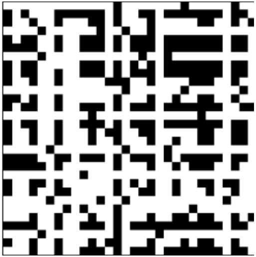
Marginal Conditional on u and v only: cMMSB^{uv} - Diagrammatic representation

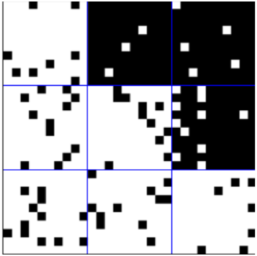


Where do we apply Copula DPs?

- The model itself may be applied in various applications. For us, we applied in a setting where
  - In Community learning (or detection) where we do NOT fix the maximum number of communities available.
  - Under many social network settings, **certain known subgroups of people** may have higher correlations in terms of their membership categories towards each other.
- Again, some introduction on Infinite relational model and infinite MMSB is needed.

- Community learning is emerging topic applicable to many social networking problems
- **hot** in machine learning.
- Partition a network of nodes into different groups based on their pairwise and directional observations (often binary)
- Data is **directional** i.e., I followed you doesn't mean you followed me.





Simple Stochastic Block Model assumption: Fixed K communities

The model:

- There is a hidden "compatibility" matrix $\mathbf{B}$, size $K \times K$, each element $\mathbf{B}_{kl} \sim Beta(\lambda_1, \lambda_2)$, a realization example:

0.5	0.2	0.1	0.1	0	...	0.1
0.3	0.91	0.2	0.4	0.2	...	0.5
...	...	...	...	...	...	0.2
0.32	0.2	0.96	0.4	0.7	...	0.9

- Suppose that person i is in latent community 2, i.e., $z_i = 2$ and person j is in latent community 3, i.e., $z_j = 3$.
- Then $\boldsymbol{e}_{ij} \sim \text{Bernoulli}(\mathbf{B}_{z_i=2, z_j=3}) = \text{Bernoulli}(0.2)$.
- $z_i \sim \pi$: some weights of communities

Inference:

- Then, our task is to perform posterior inference on: $\Pr(z_1, \dots z_n, \pi, \mathbf{B} | \{\boldsymbol{e}_{ij}\})$

Literatures: Infinite Relational Model: K can go infinity

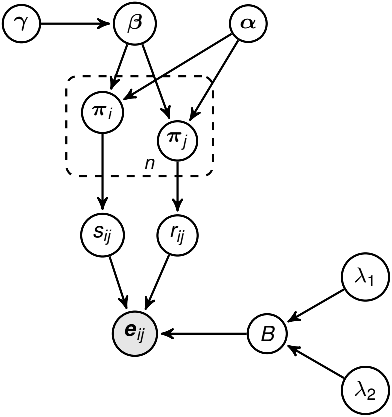
- Early work assumes a fixed number of K communities.
- However, in many applications, an accurate guess of K can be impractical.
- **Infinite Relational Model** (Kemp 2006)
- in Infinite Relational Model, K can potentially be ∞.
- **drawback** is assume each node i belong to only a single community k (i.e., $z_i = k$).

Mixed-Membership Stocastic Blockmodel (MMSB)
Airoldi, Blei, Fienberg, Xing (2008)

- **mixed-membership concept** each node i may belong to multiple communities.
- each node i has its own community distribution π_i, like topic distribution in a document (LDA).
- hence each node has its own **edge** membership indictors $\{s_{i\cdot}, r_{\cdot i}\}$
- e_{ij} are sampled using pairs of membership indicators (s_{ij}, r_{ji}).

Generative Model (of the infinite version)

1. $\boldsymbol{\beta} \sim GEM(\gamma)$
2. $\{\pi_i\}_{i=1}^{n} \sim DP(\alpha \cdot \boldsymbol{\beta})$
3. $s_{ij} = \pi_i, \;\; r_{ij} = \pi_j$
4. $B_{k,l} \sim Beta(\lambda_1, \lambda_2), \forall k, l;$
5. $\boldsymbol{e}_{ij} \sim Bernoulli(B_{s_{ij}, r_{ij}})$.



The priors:

- Each element $\mathbf{B}_{kl} \sim Beta(\lambda_1, \lambda_2)$ still: a realization example:

0.5	0.2	0.1	0.1	0	...	0.1
0.3	0.91	0.2	0.4	0.2	...	0.5
...	...	...	...	...	...	0.2
0.32	0.2	0.96	0.4	0.7	...	0.9

- Suppose that interaction i **sent to** j is of latent community 2, i.e., $s_{ij} = 2$,
- Interaction j **received from** i is in latent community 3, i.e., $r_{ji} = 3$.
- Note that s_{ij} do not generally equal r_{ji}.
- Then $\boldsymbol{e}_{ij} \sim \text{Bernoulli}(\mathbf{B}_{s_{ij}=2, r_{ji}=3}) = \text{Bernoulli}(0.2)$.
- $\{s_{i,k}, r_{i,k}\} \sim \pi_i$: There are altogether N πs

The posterior

- Then, our task is to perform posterior inference on:
  $\Pr(\{s_{i,j}, r_{j,i}\}_{\forall, 1 \leq i,j \leq N}, \mathbf{B}, \pi_1, \ldots \pi_N | \{\boldsymbol{e}_{ij}\})$

Recent variants of MMSB

A few variants were subsequently proposed from MMSB, examples include:

- [**?**] extends the mixture-membership model with a dynamic setting;
- [**?**] extends the MMSB into the infinite case; and
- [**?**] incorporates the node's metadata information into MMSB.

Our work: Copula Mixed-Membership Stochastic Blockmodel with Subgroup Correlation

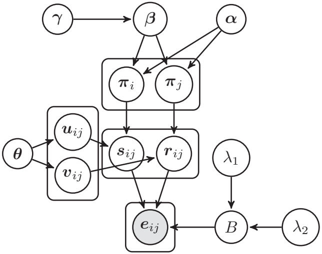
- Despite MMSB's powerful representations,it assumes distributions of community membership indicators between the two nodes are independent.
- Under many social network settings, **certain known subgroups of people** may have higher correlations in terms of their membership categories towards each other.
- We introduce a new framework where individual Copula function is to be employed to model jointly the membership pairs of those nodes within the subgroup of interest.
- Various Copula functions may be used to suit the scenario, while maintaining the membership's marginal distribution, as needed for modeling membership indicators with other nodes outside of the subgroup of interest.
- Experimental results shows a superior performance when comparing with the existing models on both the synthetic and real world datasets.

Copula DP is added to the model

Generative Model

- C1: $\boldsymbol{\beta} \sim GEM(\gamma)$
- C2: $\{\pi_i\}_{i=1}^{n} \sim DP(\alpha \cdot \boldsymbol{\beta})$
- C3: $\begin{cases} (u_{ij}, v_{ij}) \sim Copula(\theta), & g_{ij} = 1; \\ u_{ij}, v_{ij} \sim U(0,1), & g_{ij} = 0. \end{cases}$
- C4: $s_{ij} = \Pi_i^{-1}(u_{ij}), r_{ij} = \Pi_j^{-1}(v_{ij})$
- C5: $B_{k,l} \sim Beta(\lambda_1, \lambda_2), \forall k, l;$
- C6: $\boldsymbol{e}_{ij} \sim Bernoulli(B_{s_{ij}, r_{ij}})$.

Graphical model



$$\Pi_i^{-1}(u_{ij}) = \{\min q : \sum_{q=1}^{k} \pi_{iq} \geq u_{ij}\}$$

$$\Pi_j^{-1}(v_{ij}) = \{\min q : \sum_{q=1}^{k} \pi_{jq} \geq v_{ij}\}$$

- Marginal Conditional on π only: after added a likelihood $\Longrightarrow$ cMMSB^{π}
- Marginal Conditional on u and v only: after added a likelihood $\Longrightarrow$ cMMSB^{uv}

Marginal Conditional on u and v only: cMMSB^{uv}

- After added the likelihood function:

$$\begin{aligned} \Pr(s_{ij} = k, r_{ij} = l | \_\_) &\propto \Pr(s_{ij} = k | u_{ij}, \boldsymbol{\beta}, \{N_{ik}\}_k) \cdot \Pr(r_{ij} = l | v_{ij}, \boldsymbol{\beta}, \{N_{jk}\}_k) \cdot \Pr(e_{ij} | \lambda_1, \lambda_2, m_{k,l}) \\ \text{where} \qquad & N_{ik} = \#\{j : s_{ij} = k\} + \#\{j : r_{j\boldsymbol{i}} = k\} \end{aligned}$$

- Each of the priors is in the form of $\Pr(s_{ij} = k | u_{ij}, \lambda_1, \dots, \lambda_K)$, we have:

$$\Pr\left(\sum_{d=1}^{k-1} \pi_{id} \leq u_{ij} \bigcap \sum_{d=1}^{k} \pi_{id} > u_{ij}\right) = \text{Betacdf}_{u_{ij}}(h_i^{k-1}, \hat{h}_i^{k-1}) - \text{Betacdf}_{u_{ij}}(h_i^{k}, \hat{h}_i^{k})$$

- In this context $\lambda_d = \alpha\beta_d + N_{id}^{-ij}$, this is because of the Dirichlet-Multinomial conjugacy.

Inference other variable of Interest

- In cMMSB^{π}, variables of interest are $\{\pi_i\}, \{s_{ij}, r_{ij}\}, \boldsymbol{\beta}$.
- In cMMSB^{uv}, variables of interestst include $\{u_{ij}, v_{ij}\}, \{s_{ij}, r_{ij}\}, \boldsymbol{\beta}$, and an auxiliary variable $\boldsymbol{m}$.
- The rest of the inference is largely similar to Hierarchical Dirichlet Process by Teh., et., all.

- When a Copula is introduced, $p(\pi_i)$ and $\Pr(s_{ij}|\pi_i)$ are no longer a conjugate pair.
- Therefore, resort to the use of Metropolis-Hastings (M-H) Sampling in each (τ)-th Gibbs iteration

For each node i, posterior distribution of π_i is:

$$
\begin{aligned}
& p(\pi_i|\alpha, \boldsymbol{\beta}, \{s_{ij}, r_{ij}\}_{i,j}) \\
\propto & \prod_{k=1}^{K+1} \pi_{ik}^{\alpha\boldsymbol{\beta}_k - 1} \cdot \prod_{j=1}^{n} \left[p_{ij}^{s_{ij} r_{ij}}(\pi_i, \pi_j) p_{ji}^{s_{ji} r_{ji}}(\pi_j, \pi_i) \right]
\end{aligned}
$$

Corresponding proposal of π_i:

$$
q(\pi_i^*|\alpha, \boldsymbol{\beta}, \{s_{ij}, r_{ij}\}_{i,j}) \propto \prod_{k=1}^{K+1} [\pi_{ik}^*]^{\alpha\boldsymbol{\beta}_k + N_{ik} - 1}
$$

Acceptance ratio becomes:

$$
A(\pi_i^*, \pi_i^{(\tau)}) = \min(1, a) \tag{1}
$$

$$
a = \frac{\prod_{j=1}^{n} \left[p_{ij}^{s_{ij} r_{ij}}(\pi_i^*, \pi_j) p_{ji}^{s_{ji} r_{ji}}(\pi_j, \pi_i^*) \right]}{\prod_{j=1}^{n} \left[p_{ij}^{s_{ij} r_{ij}}(\pi_i^{(\tau)}, \pi_j) p_{ji}^{s_{ji} r_{ji}}(\pi_j, \pi_i^{(\tau)}) \right]} \cdot \frac{\prod_{k=1}^{K+1} \left[\pi_{ik}^{(\tau)} \right]^{N_{ik}}}{\prod_{k=1}^{K+1} \left[\pi_{ik}^* \right]^{N_{ik}}} \tag{2}
$$

Inference $cMMSB^{\pi}$ - Sampling $\{s_{ij}, r_{ij}\}$

$$\begin{aligned} &\Pr(s_{ij}, r_{ij} | e_{ij}, \lambda_1, \lambda_2, \theta_d, \pi_i, \pi_j, \{(s_{ij}, r_{ij})\}_{i,j}) \\ \propto & p_{ij}^{s_{ij}, r_{ij}}(\pi_i, \pi_j) \cdot p(e_{ij} | \lambda_1, \lambda_2, \{(s_{ij}, r_{ij})\}_{i,j}) \end{aligned}$$

$$p(e_{ij} | \lambda_1, \lambda_2, \{(s_{ij}, r_{ij})\}_{i,j}) = \begin{cases} n_{s_{ij}, r_{ij}}^{1, -e_{ij}} + \lambda_1, & e_{ij} = 1; \\ n_{s_{ij}, r_{ij}}^{0, -e_{ij}} + \lambda_2, & e_{ij} = 0. \end{cases}$$

Inference $cMMSB^{\pi}$ - Sampling β

- obvious choice of M-H proposal of β its prior $p(\beta|\gamma) = GEM(\gamma)$.
- this proposal can be non-informative, which results in a low acceptance rate.
- We sample β^* conditioned on an auxiliary variable $\boldsymbol{m}$: $(\beta_1^*, \cdots, \beta_K^*, \beta_{K+1}^*) \sim Dir(\boldsymbol{m}_1, \cdots, \boldsymbol{m}_K, \gamma)$, in order to increase the M-H's acceptance rate
- instead of sampling β directly from $\boldsymbol{m}$ as in [**?**], we only use it for our proposal distribution, as we have explicitly sampled $\{\pi_i\}_{i=1}^n$. The acceptance ratio is hence:

$$A(\beta^*, \beta^{(\tau)}) = \min(1, a)$$

$$a = \frac{\prod_{i=1}^n \left[\prod_{d=1}^{K+1} \Gamma(\alpha\beta_d^{(\tau)}) \cdot \pi_{id}^{\alpha\beta_d^*}\right]}{\prod_{i=1}^n \left[\prod_{d=1}^{K+1} \Gamma(\alpha\beta_d^*) \cdot \pi_{id}^{\alpha\beta_d^{(\tau)}}\right]} \cdot \frac{\prod_{d=1}^K \left[\beta_d^{(\tau)}\right]^{\boldsymbol{m}_d - \gamma}}{\prod_{d=1}^K \left[\beta_d^*\right]^{\boldsymbol{m}_d - \gamma}} \tag{3}$$

Inference $cMMSB^{u,v}$ - Sampling (u_{ij}, v_{ij})

The Copula function is used as its proposal, and therefore, its corresponding acceptance ratio becomes that of:

$$A\left((u_{ij}^{(\tau)}, v_{ij}^{(\tau)}), (u_{ij}^{*}, v_{ij}^{*})\right) = \min(1, a)$$

$$a = \frac{I_{u_{ij}^{*}}(h_i^{k-1}, \hat{h}_i^{k-1}) - I_{u_{ij}^{*}}(h_i^{k}, \hat{h}_i^{k})}{I_{u_{ij}^{(\tau)}}(h_i^{k-1}, \hat{h}_i^{k-1}) - I_{u_{ij}^{(\tau)}}(h_i^{k}, \hat{h}_i^{k})} \cdot \frac{I_{v_{ij}^{*}}(h_j^{l-1}, \hat{h}_j^{l-1}) - I_{v_{ij}^{*}}(h_j^{l}, \hat{h}_j^{l})}{I_{v_{ij}^{(\tau)}}(h_j^{l-1}, \hat{h}_j^{l-1}) - I_{v_{ij}^{(\tau)}}(h_j^{l}, \hat{h}_j^{l})}$$

Here $h_i^k, \hat{h}_i^k$'s definitions are the same as in Eq. (7) in the paper; assuming $s_{ij} = k, r_{ij} = l$.

Inference $cMMSB^{u,v}$ - Sampling s_{ij}, r_{ij}

$$
\begin{aligned}
&\Pr(s_{ij}=k, r_{ij}=l|e_{ij}, \lambda_1, \lambda_2, n_{kl}, u_{ij}, v_{ij}, \{h_i^k\}_k, \{\hat{h}_i^k\}_k, \{h_j^k\}_k, \{\hat{h}_j^k\}_k) \\
\propto &\Pr(s_{ij}=k|u_{ij},, \{h_i^k\}_k, \{\hat{h}_i^k\}_k) \cdot \Pr(r_{ij}=l|v_{ij}, \{h_j^k\}_k, \{\hat{h}_j^k\}_k) \cdot \Pr(e_{ij}|\lambda_1, \lambda_2, n_{kl}) \\
\propto &(I_{u_{ij}}(h_i^{k-1}, \hat{h}_i^{k-1}) - I_{u_{ij}}(h_i^k, \hat{h}_i^k)) \cdot (I_{v_{ij}}(h_j^{l-1}, \hat{h}_j^{l-1}) - I_{v_{ij}}(h_j^l, \hat{h}_j^l)) \cdot \Pr(e_{ij}|\lambda_1, \lambda_2, n_{kl})
\end{aligned}
$$

The likelihood is:

$$
\begin{aligned}
&\Pr(e_{ij}|s_{ij}=k, r_{ij}=l, \lambda_1, \lambda_2, n_{kl}^{-e_{ij}}) \\
\propto &P(e_{ij}, \boldsymbol{e}\backslash\{e_{ij}\}, s_{ij}=k, r_{ij}=l,, n_{kl}^{-e_{ij}}, \lambda_1, \lambda_2) \\
= &\frac{\Gamma(e_{ij}+n_{k,l}^1+\lambda_1)\Gamma(1-e_{ij}+n_{k,l}^0+\lambda_2)}{\Gamma(1+n_{k,l}+\lambda_1+\lambda_2)} \\
&P(e_{ij}|\boldsymbol{e}\backslash\{e_{ij}\}, s_{ij}=k, r_{ij}=l, \boldsymbol{Z}\backslash\{s_{ij}, r_{ij}\}, \lambda_1, \lambda_2) \\
= &\begin{cases} \frac{n_{k,l}^1+\lambda_1}{n_{k,l}+\lambda_1+\lambda_2} & \text{if } e_{ij}=1; \\ \frac{n_{k,l}^0+\lambda_2}{n_{k,l}+\lambda_1+\lambda_2} & \text{if } e_{ij}=0. \end{cases}
\end{aligned}
$$

Here $n_{k,l} = \sum_{i'j'} \mathbf{1}(s_{i'j'}=k, r_{i'j'}=l)$, $n_{k,l}^1 = \sum_{s_{i'j'}=k, r_{i'j'}=l} e_{i'j'}$, and $n_{k,l}^0 = n_{k,l} - n_{k,l}^1$.

- Selected and report the results on 3 datasets. NIPS Co-authorship dataset, the lazega-lawfirm dataset and the MIT Reality Mining dataset
- ten-folds cross-validation to complete this task, where we randomly select one out of ten for each node's link data as test data and the others as training data
- The criteria in evaluating this predict ability includes the train error ($0-1$ loss), the test error ($0-1$ loss), the test log likelihood and the AUC (Area Under the roc Curve) score obtain either comparable or better performance with other state-of-the-art.

Experiment - Result

Table: Different models' performance (Mean $\mp$ Standard Deviation) on Real world datasets

dataset		*Train error*	*Test error*	*Test log likelihood*	*AUC*
NIPS co-author	*IRM*	0.0317 ∓ 0.0004	0.0423 ∓ 0.0014	-135.0467 ∓ 7.3816	0.8901 ∓ 0.0162
	LFRM	0.0482 ∓ 0.0794	0.0239 ∓ 0.0735	-105.2166 ∓ 179.5505	0.9348 ∓ 0.1667
	MMSB	0.0132 ∓ 0.0042	0.0301 ∓ 0.0064	-86.2134 ∓ 10.1258	0.9524 ∓ 0.0215
	iMMM	$\mathbf{0.0061 \mp 0.0019}$	0.0253 ∓ 0.0035	-83.4264 ∓ 9.4293	0.9574 ∓ 0.0155
	$cMMSB^{\pi}$	0.0066 ∓ 0.0038	$\mathbf{0.0231 \mp 0.0043}$	-83.4261 ∓ 9.4280	0.9569 ∓ 0.0159
	$cMMSB^{uv}$	0.0097 ∓ 0.0047	0.0240 ∓ 0.0065	$\mathbf{-83.4257 \mp 9.4292}$	$\mathbf{0.9581 \mp 0.0153}$
MIT realtiy	*IRM*	0.0627 ∓ 0.0002	0.0665 ∓ 0.0004	-133.8037 ∓ 1.1269	0.8261 ∓ 0.0047
	LFRM	0.0397 ∓ 0.0017	0.0629 ∓ 0.0037	-143.6067 ∓ 10.0592	0.8529 ∓ 0.0179
	MMSB	0.0243 ∓ 0.0105	0.0716 ∓ 0.0043	-129.4354 ∓ 7.6549	0.8561 ∓ 0.0176
	iMMM	0.0297 ∓ 0.0055	0.0625 ∓ 0.0015	-126.7876 ∓ 3.4774	0.8617 ∓ 0.0124
	$cMMSB^{\pi}$	0.0246 ∓ 0.0016	0.0489 ∓ 0.0016	-125.3876 ∓ 3.2689	0.8794 ∓ 0.0159
	$cMMSB^{uv}$	0.0283 ∓ 0.0035	0.0438 ∓ 0.0015	-123.3876 ∓ 3.1254	0.8738 ∓ 0.0364
Lazega lawfirm	*IRM*	0.0987 ∓ 0.0003	0.1046 ∓ 0.0012	-201.7912 ∓ 3.3500	0.7056 ∓ 0.0167
	LFRM	0.0566 ∓ 0.0024	0.1051 ∓ 0.0064	-222.5924 ∓ 16.1985	0.7970 ∓ 0.0197
	MMSB	0.0391 ∓ 0.0071	0.0913 ∓ 0.0030	-212.1256 ∓ 3.2145	0.7789 ∓ 0.0102
	iMMM	0.0487 ∓ 0.0068	0.1096 ∓ 0.0026	-202.7148 ∓ 5.3076	0.7874 ∓ 0.0141
	$cMMSB^{\pi}$	0.0246 ∓ 0.0050	0.1023 ∓ 0.0056	-201.0154 ∓ 5.2167	0.8273 ∓ 0.0148
	$cMMSB^{uv}$	0.0276 ∓ 0.0043	0.1143 ∓ 0.0019	-204.0289 ∓ 9.5460	0.8215 ∓ 0.0167