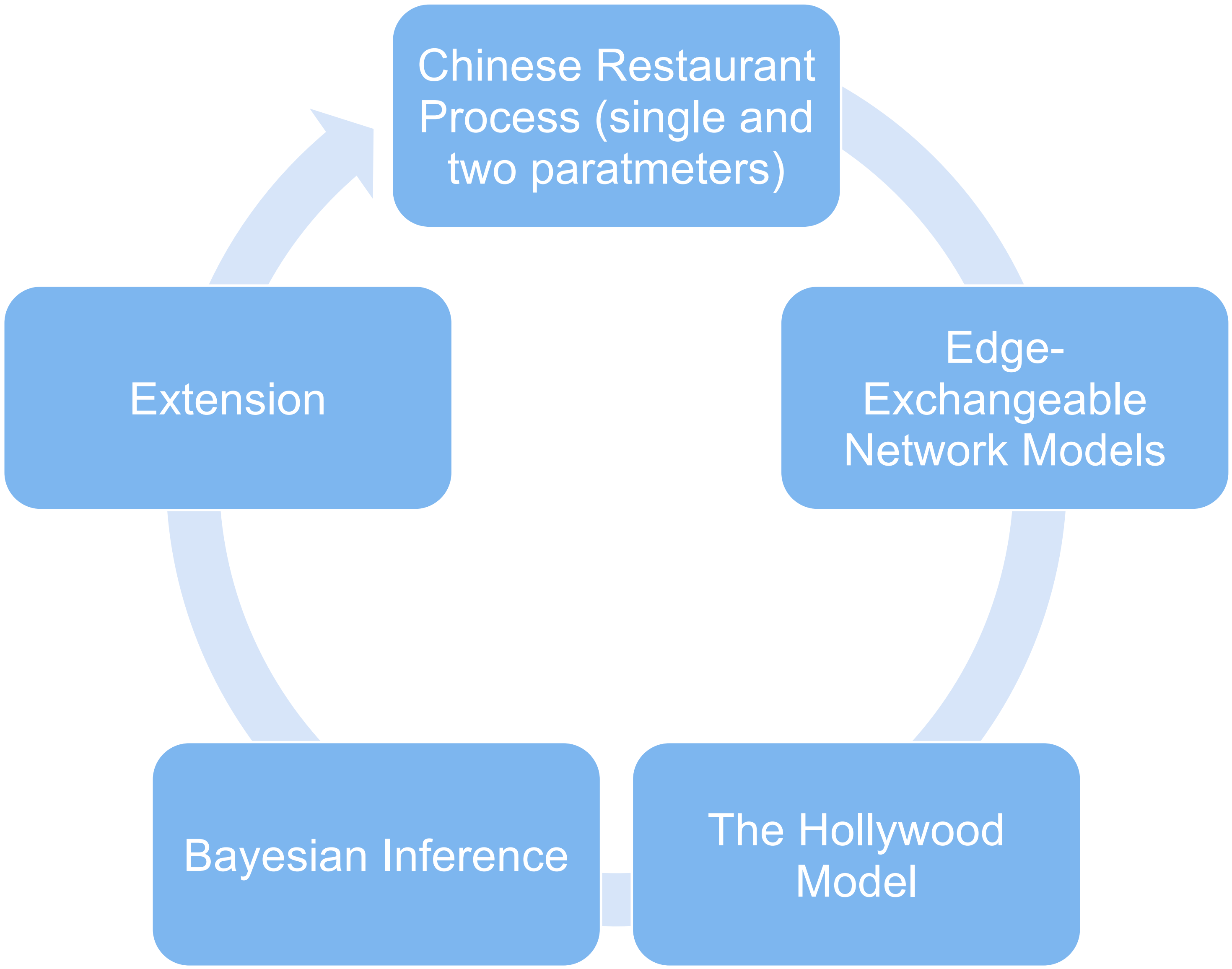
Edge-exchangeable random graph models

Xiaoyu Ouyang

2021/03/10

Outline


Chinese Restaurant Process (single and two paratmeters)
Edge-Exchangeable Network Models
The Hollywood Model
Bayesian Inference
Extension

Chinese Restaurant Process

Chinese Restaurant Process: a generative process for a certain class of exchangeable partitions which satisfies the following two properties:

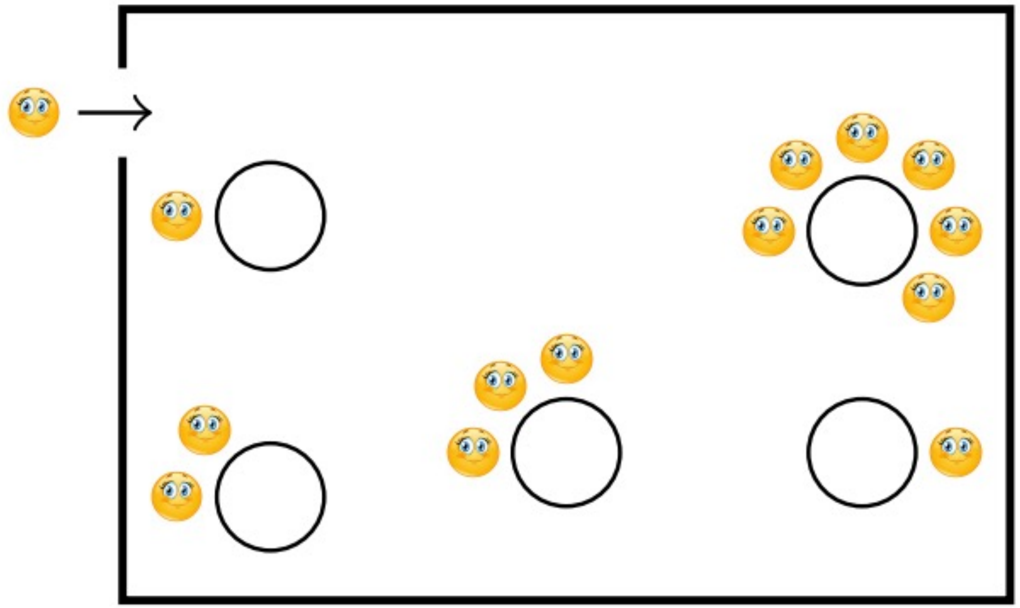
- Exchangeable random partition

For any n, the distribution is invariant w.r.t. any permutation of [n]

- The probability of creating a new cluster only depends on the sample size n

$$P(C^{n+1} = \text{new} \mid C^1, C^2, \ldots, C^n) = f(n)$$

One-Parameter Chinese Restaurant Process

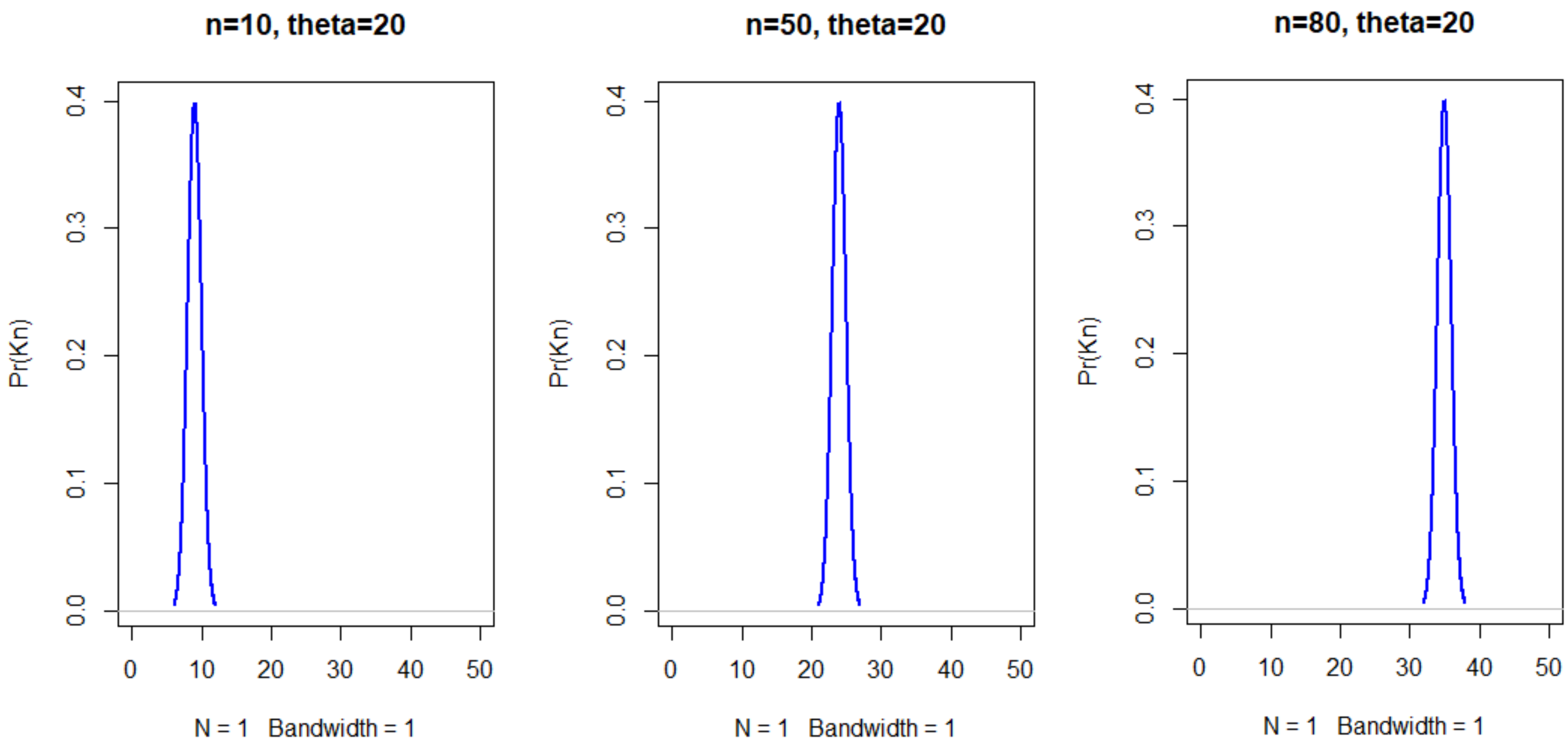


Imagine a restaurant with countably infinitely many tables, labelled 1, 2, ...
Customers walk in and sit down at some table.

- The 1st customer always chooses the first table.

- Customer n + 1
  - Joins an existing table j = 1, . . . , K, w.p. $\frac{m_{n,j}}{n+\vartheta}$
  - Sits at a new table w.p. $\frac{\vartheta}{n+\vartheta}$

One-Parameter Chinese Restaurant Process



Two-Parameter Chinese Restaurant Process

- CRP (α, ϑ, n)
  - The 1st customer always chooses the first table

  - Customer n+1:
    - Join an existing table k=1,...,K_n, w.p. $\frac{m_{n,k}-\alpha}{n+\vartheta}$
    - Sits at a new table w.p. $\frac{K_n * \alpha+\vartheta}{n+\vartheta}$

- Two parameters: $0 \leq \alpha < 1 \quad \vartheta > -\alpha$
- $\alpha = 0$: One parameter CRP
- Simulated data e.g. (ϑ=2, α=0.5, n=20)

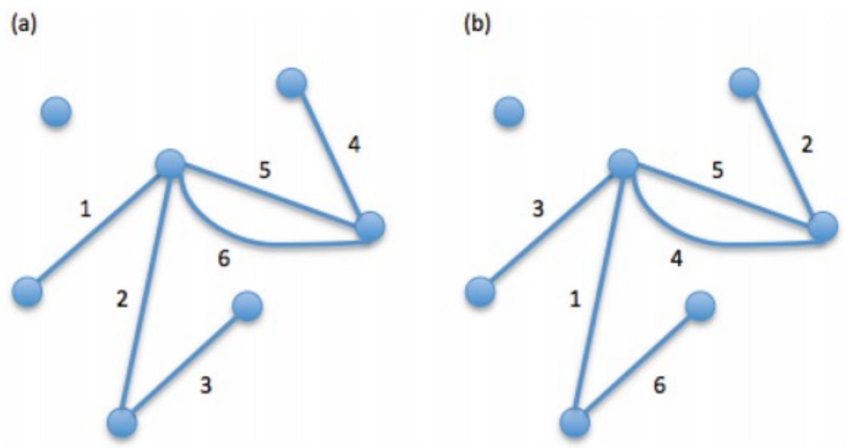
```
> CRP2(2,0.5,20)
 [1] 1 2 3 4 5 6 3 5 5 4 3 3 3 3 5 3 3 5 3 7
```

Edge-exchangeable Network Models

- Y: a random edge-labeled network
- S: edges labeled in a coutable set
- σ : a permutation from S to S
- Y^{σ}: the edge-labeled network induced by $I^{\sigma} : S \to \text{fin}(P)$ with
- $I^{\sigma}(i) = I(\sigma^{-1}(i))$ for each $i \in S$.

➢ A random edge-labeled network Y is edge-exchangeable if $Y^{\sigma} = Y$ for all permutations σ, where = denotes equality in distribution.

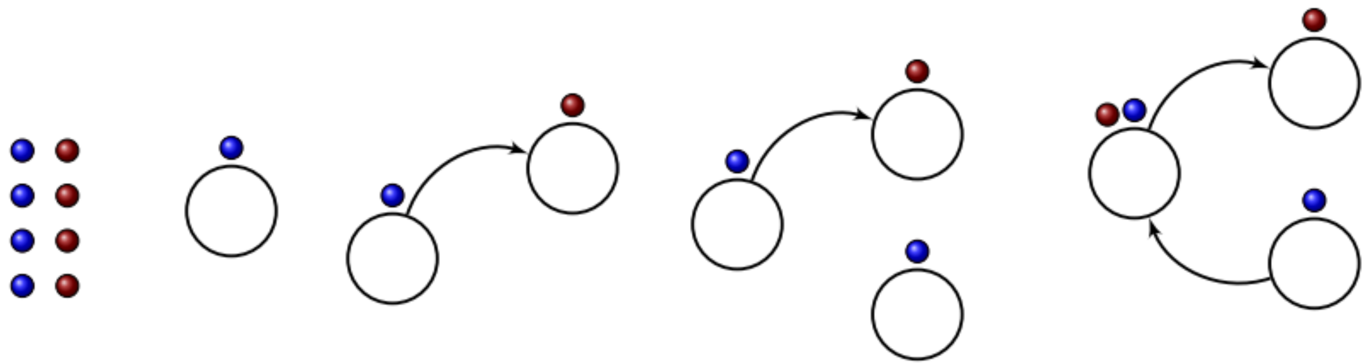
➢ Edge exchangeable models assign the same probability to all edge-labeled networks that are isomorphic up to relabeling.



The Hollywood Model

- A process for sampling the end-points of the edges
- Table/cluster = nodes; Customer = half-edge
- For n edges, 2n steps of a Chinese restaurant process
- Here only consider the binary Hollywood Process

- Based on the 2-parameter Chinese Restaurant process
  - K number of nodes with at least one connection, with degree $m_j \geq 1$
  - Pick a node as starting/end node with probability prop. to $m_j - \alpha$
  - Choose a new node with probability prop. to $\vartheta + \alpha K$



Interpretation of parameters

- $0 \leq \alpha < 1$; $\vartheta > -\alpha$

I. Sparsity $\alpha \in (1/2, 1)$; Power-law degree distribution

II. $(\alpha > 0)$ $\alpha \uparrow$: prob of observing unseen vertices $\uparrow$

prob of seeing a vertex again $\downarrow$

III. $\vartheta \uparrow$: prob of observing unseen vertices $\uparrow$

effect of ϑ diminishes as n tends to infinity

- $\alpha < 0$; $\vartheta = -k\alpha$ for $k = 1,2,3...$ (finite number of vertices)

I. Opposite effects

II. $v(\varepsilon) \to k$ as $n \to \infty$

Inference

- Oriented network likelihood calculated as the product of conditional densities:

$$P(Y_n = \varepsilon; \alpha, \vartheta) = a^{v(\varepsilon)} \frac{(\vartheta/a)^{\uparrow v(\varepsilon)}}{\vartheta^{\uparrow m(\varepsilon)}} \prod_{k=2}^{\infty} (exp(N_k(\varepsilon))(1-a)^{\uparrow(k-1)})$$

- $v(\varepsilon)$: number of non-isolated vertices
- $N_k(\varepsilon)$: number of vertices with degree k>=1
- $m(\varepsilon)$: total degree of ε
- $x^{\uparrow j} = x(x+1)(x+2)\ldots(x+j-1)$

Log-likelihood is calculated as the following:

- $l(\alpha, \vartheta; \varepsilon) = v(\varepsilon) log(a) + \sum_{i=0}^{v(\varepsilon)-1} log(\frac{\vartheta}{a} + j) - \sum_{j=1}^{m(\varepsilon)} log(\vartheta + j - 1) + \sum_{k=2}^{\infty} \sum_{j=0}^{k-2} log(1 - a +$

```
> log.lik(a,1000,0.2,3)
[1] -12782.28
```

Maximum Likelihood Estimate

- Partial derivatices are set to 0

$$\frac{\partial l\ (\alpha,\vartheta,;\varepsilon)}{\partial \alpha} = \frac{v(\varepsilon)}{a} + \sum_{j=0}^{v(\varepsilon)-1} \frac{-\vartheta/a^2}{\vartheta/a+j} - \sum_{k=2}^{\infty}\sum_{j=0}^{k-2} \frac{N_k(\varepsilon)}{1-a+j}$$

$$\frac{\partial l\ (\alpha,\vartheta,;\varepsilon)}{\partial \vartheta} = \sum_{j=0}^{v(\varepsilon)-1} \frac{1/a}{\vartheta/a+j} - \sum_{j=0}^{m(\varepsilon)-1} \frac{1}{\vartheta+j}$$

- Confidence intervals
- Quasi-Newton Method： confine to $0 < a < 1$ or $a < 0$
- $a = 0$: $\vartheta_{mle} = \frac{v(\varepsilon)}{log(m(\varepsilon))}$
- CLT: $\sqrt{log(2n)}\,(\vartheta_{mle} - \vartheta) \to N(0,1)$ $as\ n \to \infty$

Comparison With Other Estimates

- Estimate minimising the Bayes risk (square error): posterior mean
- Prior $\quad a \sim U[0,1];\ \vartheta \sim Gamma(10, 0.5)$ independent

  Likelihood $\quad P(Y_n = \varepsilon; \alpha, \vartheta)$

  Posterior $\quad Q \propto \pi(a)\pi(\vartheta)P(Y_n = \varepsilon; \alpha, \vartheta)$ (Target Distribution)

- Metropolis-Hastings Algorithm with proposal distribution q(. |):

Start with $a^{(0)}, \vartheta^{(0)}$, for t=1,2,3:

1. Sample $(a^*, \vartheta^*) \sim q(\cdot\, | a^{t-1}, \vartheta^{t-1})$
2. Compute acceptance probability:

$$P(a^*, \vartheta^* | a^{t-1}, \vartheta^{t-1}) = min(1, \frac{Q(a^*, \vartheta^*)q(a^{t-1}, \vartheta^{t-1} | a^*, \vartheta^*)}{Q(a^{t-1}, \vartheta^{t-1})q(a^*, \vartheta^* | a^{t-1}, \vartheta^{t-1})})$$

3. Sample $U \sim U[0,1]$. If $U \leq P(a^*, \vartheta^* | a^{t-1}, \vartheta^{t-1})$, set $a^{(t)} = a^*, \vartheta^{(t)} = \vartheta^*$; otherwise set $a^{(t)} = a^{(t-1)}, \vartheta^{(t)} = \vartheta^{(t-1)}$

- Calculate posterior mean and compare with MLE and other estimates using

More Structured Models (Extension)

Relax the the exchangeability assumption

- The Hollywood model:

$$e_n = (e_{n1}, e_{n2}) = (starting\ half-edge, ending\ half-edge)$$

Given $v_1, v_2, \ldots, v_{2n-1}, v_{2n}$, we are able to consturct a graph with n edges.

Generative process:

$$P(v_k = j) = \frac{m_j - \alpha}{k - 1 + \vartheta}$$

$$P(v_k = new) = \frac{K_k * \alpha + \vartheta}{k - 1 + \vartheta}$$

- Change $m_j = \sum_{i=1}^{k-1} I\{v_i = j\}$ to break the exchangeability assumption:

$$m_j = \sum_{i=k-1-t}^{k-1} I\{v_i = j\}$$

The probability $P(v_k = j)$ is now dependent on the last t vertices

Inference of the extended model

- Log-likelihoods, Maximum Likelihood Estimates, Confidence Intervals

- Other properties

Conclusion

The Hollywood model

- computationally tractable
- performs well in many real-data examples
- nice properties such as sparsity and power-law degree distribution

The extended model

- better fit to real-life data?
- nice properties?

Thanks for listening!