Posterior Distribution of a Random Process Conditioned on Empirical Frequencies of a Finite Path

The i.i.d and finite Markov chain case

Wenqing Hu [1]

[1]Department of Mathematics and Statistics, Missouri S&T. Presentation based on joint work with Hong Qian (Department of Applied Mathematics, University of Washington). Preprint avilable at arXiv :2202.11780 [math.PR]

Axioms of Probability Theory

- Axioms of Probability Theory (A.N.Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, 1933) : $(\Omega, \mathcal{F}, \mathbf{P})$
- Ω is a sample space of elementary events, $\mathcal{F}$ is a σ-algebra on the sample space, and $\mathbf{P}$ is a probability measure given a-priori.
- Usually one does not know $\mathbf{P}$ in any realistic way. However, $\mathbf{P}$ has to be given first before probabiliy calculations can be carried on.

Statistical Inference

- *"Statistical inferences are based only in part upon the observations. An equally important base is formed by prior assumptions about the underlying situation. Even in the simplest cases, there are explicit or implicit assumptions about randomness and independence, about distributional models, perhaps prior distributions for some unknown parameters, and so on."*

  – P.J.Huber, *Robust Statistics.* John Wiley & Sons, 1981

Building Statistics from Probability and Vice Versa

- The route of building statistical models from probability, as an integral part of data science, thus, should always start with the "basic assumption of a probability space including a prior probability measure".
- Reverse Direction : **Building Probability from Statistics ?**
- Bayesian logic. Frequentist ?
- Development in the Ergodic Theory and Dynamical Systems. (Sinai, Ornstein, etc.)

Our question

- **Our question :** *Given an observation of the empirical frequencies of a random process, to what extent can we recover the probability structure of the original random process via conditioning ?*
- A very simple but attractive problem !

Two Paradigms of Doing Science : Newtonian an Keplerian

- Two different paradigms of doing scientific research [2].
- The Newtonian paradigm : the first-principle-based approach.
- The Keplerian paradigm : the data-driven approach.
- The current Data Science is in the Keplerian Paradigm !

[2] Refer to W. E., The dawning of a new era in applied mathematics. *Notices of the American Mathematical Society*, **68**(4) :565-571, 2021.

The i.i.d case : Set-up

- Let X_1, ..., X_n, ... be an i.i.d sequence defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with common distribution as a random variable X taking values in $\mathbb{N}$.
- Given a sequence of sample frequencies $\nu_k \in \mathbb{N}_+$ satisfying

$$\sum_{k\in\mathbb{N}} \nu_k = n \ ,$$

we consider the event

$$\mathcal{E}_{\{\nu_k\}} = \left\{ \sum_{\ell=1}^{n} \mathbb{1}_k(X_\ell) = \nu_k, k \in \mathbb{N} \right\},$$

where $\mathbb{1}_k(X_\ell) = \begin{cases} 1 \ , & \text{if } X_\ell = k \ , \\ 0 \ , & \text{otherwise} \ . \end{cases}$

The i.i.d case : Set-up

- $\mathcal{E}_{\{\nu_k\}}$ stands for the event that the trajectory $X_\ell, \ell = 1, 2, ..., n$ takes on value k with frequency ν_k, $k \in \mathbb{N}$, respectively.
- Example : $X_1 = 1, X_2 = 2, X_3 = 1, X_4 = 1, X_5 = 4$. Then $\mathcal{E}_{\{\nu_k\}} = \{\nu_1 = 3, \nu_2 = \nu_4 = 1, \nu_k = 0 \text{ for all other } k\}$.
- Given $\mathcal{E}_{\{\nu_k\}}$, what is the posterior distribution of X_ℓ where $\ell = 1, 2, ..., n$?

The i.i.d case : Theorem

Theorem (posterior distribution for the i.i.d. case)

Given $m \in \mathbb{N}$ and any $1 \leq \ell \leq n$, we have

$$\mathbf{P}\big(X_\ell = m \big| \mathcal{E}_{\{\nu_k\}}\big) = \frac{\nu_m}{n} \ .$$

The i.i.d case : Heuristics

- Why ? Simple symmetry leads to the fact.
- Think of

Case	X_1	X_2	X_3	X_4	X_5
a	1	2	1	1	4
b	2	1	1	4	1
c	1	1	1	2	4

- In each case, the joint probability is the same !
- **Conditional Symmetry** : As long as $\mathcal{E}_{\{\nu_k\}}$ is observed, the joint probability of $X_1, \ldots, X_n$ is symmetric with respect to any permutations of the outcomes.

The i.i.d case : Heuristics

- Think of one outcome

$$(1,2,1,1,4)\begin{cases}(\color{red}{1},\color{orange}{2},\color{blue}{1},\color{green}{1},\color{purple}{4})\\(\color{red}{1},\color{orange}{2},\color{green}{1},\color{blue}{1},\color{purple}{4})\\(\color{blue}{1},\color{orange}{2},\color{red}{1},\color{green}{1},\color{purple}{4})\\(\color{blue}{1},\color{orange}{2},\color{green}{1},\color{red}{1},\color{purple}{4})\\(\color{green}{1},\color{orange}{2},\color{blue}{1},\color{red}{1},\color{purple}{4})\\(\color{green}{1},\color{orange}{2},\color{red}{1},\color{blue}{1},\color{purple}{4})\end{cases}$$

- We can color the same outcome 1 with different colors to create "different" outcome sequences.
- We "lift" the conditional probability $\mathbf{P}(\bullet|\mathcal{E}_{\{\nu_k\}})$ to a new probability $\mathscr{P}$ on the space of colored sequences with given frequency event $\mathcal{E}_{\{\nu_k\}}$.

The i.i.d case : Heuristics

- In the space of colored sequences any probability

$$\begin{aligned} & \mathscr{P}(X_k = \text{some colored } m) \\ \propto \quad & \text{Number of colored trajectories such that} \\ & \qquad X_k = \text{some colored } m \end{aligned}$$

- A higher level of **Conditional Symmetry** : This number is the same no matter how you choose the colored m, as long as the frequency event $\mathcal{E}_{\{\nu_k\}}$ is given.

The i.i.d case : Soft proof

- So actually $\mathscr{P}(X_k = \text{some colored } m) = \dfrac{1}{n}$.
- And thus
  $\mathbf{P}(X_k = m | \mathcal{E}_{\{\nu_k\}}) = \sum_{i=1}^{\nu_m} \mathscr{P}(X_k = m \text{ with color } i) = \dfrac{\nu_m}{n}$.
- The Theorem is softly proved !

The i.i.d case : Conditional Symmetry ideas

- Two levels of conditional symmetries are used in the i.i.d case.
- Conditional Symmetry at the level of sample path trajectories : As long as $\mathcal{E}_{\{\nu_k\}}$ is given, we can permute any of the realizations of $(X_1, ..., X_n)$ in a trajectory without changing the joint probability [3].
- Conditional Symmetry at the level of individual observations : As long as $\mathcal{E}_{\{\nu_k\}}$ is given, fix $X_k =$ some colored m, then the number of colored trajectories such that $X_k =$ some colored m is independent of the colored m.

---

[3] It is related to de Finetti's "exchangeablilty", see P. Diaconis and D. Freedman, de Finettis theorem for Markov chains. *The Annals of Probability*, **8** :115-130, 1980. In that work the authors are considering another problem, that exchangebility implies some "averaged" transition probability.

The i.i.d case : Empirical Frequency as Posterior Probability

$$\text{Conditional Symmetry} \overset{\text{leads to}}{\longrightarrow} \begin{array}{l}\text{Empirical Frequency} \\ = \text{Posterior Probability}.\end{array}$$

The finite Markov chain case : Set-up

- Is the above a general philosophy ?
- Finite Markov chain case : $Y_1, ..., Y_n, ...$ is a time-homogeneous Markov chain with finite state space $\Sigma = \{1, ..., N\}$, $|\Sigma| = N$.
- Transition probability matrix is $P = (p_{ij})_{1 \leq i,j \leq N}$.
- Assume the process starts from an initial probability distribution $\pi^0 = (\pi_1^0, ..., \pi_N^0)$, $0 \leq \pi_i^0 \leq 1$, $\sum_{i=1}^{N} \pi_i^0 = 1$, such that $\mathbf{P}(Y_1 = i) = \pi_i^0$.

The finite Markov chain case : Set-up

- How do we count "empirical frequencies" in this case ?
- Consider the "consecutive pair" process $X_\ell = (Y_\ell, Y_{\ell+1}), \ell \geq 1$.
- Given a sequence of sample frequencies $\nu_{(i,j)} \in \mathbb{N}_+$ satisfying

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \nu_{(i,j)} = n \ ,$$

we consider the event

$$\mathcal{E}_{\{\nu_{(i,j)}\}} = \left\{ \sum_{\ell=1}^{n} \mathbb{1}_{(i,j)}(X_\ell) = \nu_{(i,j)}, 1 \leq i,j \leq N \right\},$$

where $\mathbb{1}_{(i,j)}(X_\ell) = \begin{cases} 1 \ , & \text{if } X_\ell = (i,j) \ , \\ 0 \ , & \text{otherwise} \ . \end{cases}$

The finite Markov chain case : Set-up

- $\mathcal{E}_{\{\nu_{(i,j)}\}}$ stands for the event that the trajectory X_ℓ $(\ell = 1, ..., n)$ takes on value (i, j) with frequency $\nu_{(i,j)}$, $1 \leq i, j \leq N$, respectively.
- Example : $(Y_1, Y_2, Y_3, Y_4, Y_5) = (1, 2, 1, 1, 2)$, then $X_1 = (1, 2), X_2 = (2, 1), X_3 = (1, 1), X_4 = (1, 2)$ and

$$\begin{aligned} \mathcal{E}_{\{\nu_{(i,j)}\}} \quad &= \{\nu_{(1,2)} = 2, \nu_{(2,1)} = \nu_{(1,1)} = 1, \\ &\qquad \nu_{(i,j)} = 0 \text{ for any other } (i,j)\} \ . \end{aligned}$$

The finite Markov chain case : Conditional Symmetry

- Do we have the same conditional symmetry as the i.i.d. case?
- Given a path $X_1 = (i_{11}, i_{12}), X_2 = (i_{21}, i_{22}), ..., X_n = (i_{n1}, i_{n2})$, the joint probability is given by

$$\begin{aligned} & \mathbf{P}\left(X_1 = (i_{11}, i_{12}), X_2 = (i_{21}, i_{22}), ..., X_n = (i_{n1}, i_{n2})\right) \\ = \quad & \pi_0^{i_{11}} p_{i_{11} i_{12}} ... p_{i_{n1} i_{n2}} \ . \end{aligned}$$

- The path has to be of a "chain type string", i.e., $i_{12} = i_{21}, ...$, etc. Based on this, one can check that given $\mathcal{E}_{\{\nu_{(i,j)}\}}$ we must have

$$\sum_{j=1}^{N} \nu_{(i,j)} - \sum_{j=1}^{N} \nu_{(j,i)} = \mathbf{1}_{\{i=i_1\}} - \mathbf{1}_{\{i=i_{n+1}\}} \ .$$

The finite Markov chain case : Conditional Symmetry

- Permutations on the entries of X_ℓ's will not change this joint probability, as long as the permutation does not "break" the chain.
- So we still have conditional symmetry at the level of sample path trajectories.

The finite Markov chain case : Constraints made by the frequency event $\mathcal{E}_{\{\nu_{(i,j)}\}}$

- For the chain case, given $\mathcal{E}_{\{\nu_{(i,j)}\}}$, the choices of Y_ℓ and X_ℓ are not arbitrary.
- Example : Suppose $\{Y_\ell\}_{\ell \geq 1}$ is a stationary Markov chain with a 3-element state space $\{1, 2, 3\}$ and stationary measure $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Set $n = 2$ and suppose we have observed

  $$\mathcal{E}_{\{\nu_{(i,j)}\}} = \left\{\nu_{(1,2)} = \nu_{(2,3)} = 1, \nu_{(i,j)} = 0 \text{ for all other pairs of } (i,j)\right\} .$$

  Then it is easy to see that $\mathbf{P}(Y_1 = 1 | \mathcal{E}_{\{\nu_{(i,j)}\}}) = 1$ while $\mathbf{P}(Y_1 = 2 | \mathcal{E}_{\{\nu_{(i,j)}\}}) = \mathbf{P}(Y_1 = 3 | \mathcal{E}_{\{\nu_{(i,j)}\}}) = 0$.

The finite Markov chain case : New Definitions

- Define $\Sigma^{\checkmark}(\ell|\mathcal{E}_{\{\nu_{(i,j)}\}}) = \left\{ i : 1 \leq i \leq N \ , \ \mathbf{P}(Y_\ell = i|\mathcal{E}_{\{\nu_{(i,j)}\}}) > 0 \right\} \ , \ 1 \leq \ell \leq n \ .$
- Define
$$\mathbf{1}_\ell^{i,\checkmark} \equiv \mathbf{1}_{\Sigma^{\checkmark}(\ell|\mathcal{E}_{\{\nu_{(i,j)}\}})}(i) \ .$$
- Given an (i,j) such that $\nu_{(i,j)} \geq 1$ on the event $\mathcal{E}_{\{\nu_{(i,j)}\}}$, we define by $\#_\ell^{(i,j)}(\mathcal{E}_{\{\nu_{(i,j)}\}})$ to be the number of different strings of chain type $X_1 = (i_1, i_2), \ldots, X_n = (i_{n-1}, i_n)$ with the ℓ-th element being $X_\ell = (i,j)$, and satisfying $\mathcal{E}_{\{\nu_{(i,j)}\}}$.

The finite Markov chain case : Theorem

Theorem (posterior distribution for the finite Markov chain case)

Given $1 \leq i, j \leq N$, *then we have*

$$\mathbf{P}\big(X_1 = (i,j) \big| \mathcal{E}_{\{\nu_{(i,j)}\}}\big) = \frac{\mathbf{1}_1^{i,\checkmark} \pi_i^0}{\sum_{k_1=1}^N \mathbf{1}_1^{k_1,\checkmark} \pi_{k_1}^0} \cdot \frac{\mathbf{1}_2^{j,\checkmark} \cdot \#_1^{(i,j)}(\mathcal{E}_{\{\nu_{(i,j)}\}})}{\sum_{k_2=1}^N \mathbf{1}_2^{k_2,\checkmark} \cdot \#_1^{(i,k_2)}(\mathcal{E}_{\{\nu_{(i,j)}\}})} .$$

The finite Markov chain case : Soft proof again

- Can be proved in a similar way as the i.i.d. case using conditional symmetry at the level of sample path trajectories.
- As before we consider the colored sequences, so something like

$$(1,2),(2,1),(1,1),(1,2)\left\{\begin{array}{l}\color{red}{(1,2)},\color{orange}{(2,1)},\color{green}{(1,1)},\color{blue}{(1,2)}\\ \color{blue}{(1,2)},\color{orange}{(2,1)},\color{green}{(1,1)},\color{red}{(1,2)}\end{array}\right.$$

- Again we can "lift" our conditional probability $\mathbf{P}(\bullet|\mathcal{E}_{\{\nu_{(i,j)}\}})$ to the colored space into a new probability $\mathscr{P}$.
- Given the frequency event $\mathcal{E}_{\{\nu_{(i,j)}\}}$, the lifted probability $\mathscr{P}$ charges all possible strings starting from the same $Y_0 = i$ with the same probability.

The finite Markov chain case : Soft proof again

- This yields

$$\begin{aligned} & \mathbf{P}(X_1 = (i,j) | \mathcal{E}_{\{\nu_{(i,j)}\}}, Y_1 = i) \\ \propto \quad & \mathbf{1}_2^{j,\checkmark} \cdot \sum_{a=1}^{\nu_{(i,j)}} (\text{Number of colored trajectories starting} \\ & \qquad \text{from a particularly colored } (i,j)) \ . \end{aligned}$$

- From here we get

$$\mathbf{P}\big(X_1 = (i,j) \big| \mathcal{E}_{\{\nu_{(i,j)}\}}, Y_1 = i\big) = \mathbf{1}_1^{i,\checkmark} \cdot \frac{\mathbf{1}_2^{j,\checkmark} \cdot \#_1^{(i,j)}(\mathcal{E}_{\{\nu_{(i,j)}\}})}{\sum_{k_2=1}^{N} \mathbf{1}_2^{k_2,\checkmark} \cdot \#_1^{(i,k_2)}(\mathcal{E}_{\{\nu_{(i,j)}\}})} \ .$$

The finite Markov chain case : Conditional Symmetry idea

- Only one level of conditional symmetry is used in the finite Markov chain case.
- Conditional Symmetry at the level of sample path trajectories : As long as $\mathcal{E}_{\{\nu_{(i,j)}\}}$ is given, we can permute any of the realizations of $(X_1, \ldots, X_n)$ in a trajectory without changing the joint probability, as long as the resulting string still forms a chain.
- How can we get Conditional Symmetry at the level of individual observations ?

Ergodic finite Markov chain case : Asymptotic Conditional Symmetry at the level of individual observations

- Assume $p_{ij} > 0$ for all $i, j = 1, 2, ..., N$.
- **Ergodic Theorem of Markov Chains** tells us that for any $\mu > 0$ we have

$$\lim_{n\to\infty} \mathbf{P}\left(\left|\frac{\nu_{(i,j)}}{n} - \pi_i p_{ij}\right| < \mu\right) = 1 \ ,$$

  where $\pi_i, i = 1, 2, ..., N$ is the invariant measure of the Markov chain $\{Y_\ell\}_{\ell\geq 1}$ and p_{ij} are the transition probabilities.
- This means that for a typical frequency event $\mathcal{E}_{\{\nu_{(i,j)}\}}$ we must have that all $\nu_{(i,j)}$ is large as n is large.

Comparing number of possible trajectories : perturbation idea

- Let the sequence $X_1, \dots, X_n$ be long enough, i.e., n is large.
- For any two $j_1, j_2 \in \{1, 2, \dots, N\}$ and $j_1 \neq j_2$, we want to compare

$$\text{card}_1^{(i,j_1)}(\mathcal{E}_{\{\nu_{(i,j)}\}}) = \begin{array}{l}\text{Number of colored trajectories} \\ \text{starting from a particularly colored } (i, j_1)\end{array}$$

with

$$\text{card}_1^{(i,j_2)}(\mathcal{E}_{\{\nu_{(i,j)}\}}) = \begin{array}{l}\text{Number of colored trajectories} \\ \text{starting from a particularly colored } (i, j_2)\end{array}$$

Comparing number of possible trajectories : perturbation idea

- Since $\nu_{(i,j)}$ is large no matter which (i,j) you pick, the replacement of (i,j_1) by (i,j_2) (only works for the colored case !) at the start of the sequence can be viewed only as a "perturbation" to the whole configuration.
- So we expect $\mathrm{card}_1^{(i,j_1)}(\mathcal{E}_{\{\nu_{(i,j)}\}}) \approx \mathrm{card}_1^{(i,j_2)}(\mathcal{E}_{\{\nu_{(i,j)}\}})$ as n is large !
- Only at the heuristic level.
- Symmetry at the $n \to \infty$ limit since when the process reaches its invariant measure, everything will look like i.i.d. case.

Comparing number of possible trajectories : Enumerative Combinatorics

- The above idea is only a heuristic argument.
- For an exact proof, we need some results in enumerative combinatorics.
- Fix some $u, v \in \{1, 2, ..., N\}$ and consider all possible strings of chain type $X_1 = (i_1, i_2), ..., X_n = (i_n, i_{n+1})$ that satisfy the given frequency event $\mathcal{E}_{\{\nu_{(i,j)}\}}$, such that $i_1 = u, i_{n+1} = v$. The total number of such strings of chain type is denoted by $N_{uv}^{(n)}(\mathcal{E}_{\{\nu_{(i,j)}\}})$.

Comparing number of possible trajectories : Enumerative Combinatorics

Theorem (P.Whittle, 1955)

We have

$$N_{uv}^{(n)}(\mathcal{E}_{\{\nu_{(i,j)}\}}) = \frac{\prod_{i=1}^{N}\left(\sum_{j=1}^{N}\nu_{(i,j)}\right)!}{\prod_{i=1}^{N}\prod_{j=1}^{N}\nu_{(i,j)}!} F^*_{vu} ,$$

where F^*_{vu} *is the* (v,u)*-th cofactor of the matrix* F^* *and* $0! = 1$*.*
Here $F^* = (\nu^*_{ij})_{1\leq i,j\leq N}$*, where*

$$\nu^*_{ij} = \begin{cases} \mathbf{1}_{\{i=j\}} - \dfrac{\nu_{(i,j)}}{\sum_{j=1}^{N}\nu_{(i,j)}} , & \textit{if} \quad \sum_{j=1}^{N}\nu_{(i,j)} > 0 , \\ \mathbf{1}_{\{i=j\}} , & \textit{if} \quad \sum_{j=1}^{N}\nu_{(i,j)} = 0 . \end{cases}$$

Comparing number of possible trajectories : Intuitions behind Whittle's formula

- The original proof of Whittle uses complex analysis techniques such as series expansions and contour integrals.
- Billingsley (1960, *Ann. Stat.*) has provided a simple proof based on recursive relations and generating functions.
- Idea : $N_{uv}^{(n)}(\mathcal{E}_{\{\nu_{(i,j)}\}}) = \sum_{w=1}^{N} N_{wv}^{(n-1)}(\mathcal{E}_{\{\widetilde{\nu}_{(i,j)}\}})$ where $\widetilde{\nu}_{(i,j)}$ is the same as $\nu_{(i,j)}$ except that the (i,j)-th element is diminished by 1.
- Use this recursion to correspond to the algebraic relation satisfied by matrix cofactors...

The finite ergodic Markov chain case : Hard proof

- Using Whittle's formula one can prove that when n is large

$$\begin{aligned} & \text{Number of colored trajectories} \\ & \quad \text{starting from a particularly colored } (i, j_1) \\ \approx\ & \text{Number of colored trajectories} \\ & \quad \text{starting from a particularly colored } (i, j_2) \end{aligned}$$

- So in the $n \to \infty$ limit we do have Conditional Symmetry at the level of individual observations !

The finite ergodic Markov chain case : Hard proof

- "Perturbation" intuitions are reflected in calculations.
- Imagine $X_1 = (i, j)$, then the sequence $X_2 = (j, i_3), ..., X_n = (i_n, i_{n+1})$ forms a string of chain type that satisfies the frequency event $\mathcal{E}_{\{\widetilde{\nu}_{(\widetilde{i},\widetilde{j})}\}}$ with

$$\widetilde{\nu}_{(\widetilde{i},\widetilde{j})} = \begin{cases} \nu_{(\widetilde{i},\widetilde{j})} \ , & \text{if } (\widetilde{i},\widetilde{j}) \neq (i,j) \ , \\ \nu_{(\widetilde{i},\widetilde{j})} - 1 \ , & \text{if } (\widetilde{i},\widetilde{j}) = (i,j) \ . \end{cases}$$

The finite ergodic Markov chain case : Hard proof

- Apply the Whittle's formula where

$$\widetilde{\nu}^*_{\widetilde{i}\widetilde{j}} = \begin{cases} \mathbf{1}_{\{i=j\}} - \dfrac{\nu_{(i,j)} - 1}{\sum_{k=1}^{N} \nu_{(i,k)} - 1} , & \text{if} \quad \widetilde{i} = i, \widetilde{j} = j \text{ and } \sum_{k=1}^{N} \nu_{(i,k)} > 1 \ , \ \nu_{(i,j)} \geq 1 \ , \\ \mathbf{1}_{\{i=\widetilde{j}\}} - \dfrac{\nu_{(i,\widetilde{j})}}{\sum_{k=1}^{N} \nu_{(i,k)} - 1} , & \text{if} \quad \widetilde{i} = i, \widetilde{j} \neq j \text{ and } \sum_{k=1}^{N} \nu_{(i,k)} > 1 \ , \ \nu_{(i,j)} \geq 1 \ , \\ \mathbf{1}_{\{i=\widetilde{j}\}} , & \text{if} \quad \widetilde{i} = i \text{ and } \sum_{k=1}^{N} \nu_{(i,k)} = 1 \ , \ \nu_{(i,j)} = 1 \ , \\ \mathbf{1}_{\{\widetilde{i}=\widetilde{j}\}} - \dfrac{\nu_{(\widetilde{i},\widetilde{j})}}{\sum_{k=1}^{N} \nu_{(\widetilde{i},k)}} , & \text{if} \quad \widetilde{i} \neq i \text{ and } \sum_{k=1}^{N} \nu_{(\widetilde{i},k)} > 0 \ , \\ \mathbf{1}_{\{\widetilde{i}=\widetilde{j}\}} , & \text{if} \quad \widetilde{i} \neq i \text{ and } \sum_{k=1}^{N} \nu_{(\widetilde{i},k)} = 0 \ . \end{cases}$$

- For any $\varepsilon > 0$, as n is large we have $|\nu^*_{kl} - \widetilde{\nu}^*_{kl}| < \varepsilon$!

The finite ergodic Markov chain case : Result

Theorem (Asymptotic of the posterier probability)

For any $\varepsilon > 0$ small enough, there exist some $M \geq 1$ and some $n_0 = n_0(\varepsilon, M) \in \mathbb{N}$ such that for any $n \geq n_0$, there exists a family of frequency events $\mathcal{E}^{\lambda}_{\{\nu_{(i,j)}\}}, \lambda \in \widetilde{\Lambda} \subseteq \Lambda$ such that

$$\mathbf{P}\left(\bigcup_{\lambda \in \widetilde{\Lambda}} \mathcal{E}^{\lambda}_{\{\nu_{(i,j)}\}}\right) \geq 1 - \frac{\varepsilon}{M} \ ,$$

and for each frequency event $\mathcal{E}^{\lambda}_{\{\nu_{(i,j)}\}}$, $\lambda \in \widetilde{\Lambda}$, the posterior probability of X_1 conditioned on $\mathcal{E}^{\lambda}_{\{\nu_{(i,j)}\}}$ is close to the unconditioned probability of X_1, i.e. for any $1 \leq i, j \leq N$ we have

$$\left| \mathbf{P}\big(X_1 = (i,j) \big| \mathcal{E}^{\lambda}_{\{\nu_{(i,j)}\}}\big) - \frac{\mathbf{1}_1^{i,\checkmark} \pi_i^0}{\sum_{k_1=1}^{N} \mathbf{1}_1^{k_1,\checkmark} \pi_{k_1}^0} \cdot p_{ij} \right| < \varepsilon \ .$$

The “prior” probability structure of the process is not necessary !

- Think over the whole procedure...
- In the Markov chain case, do we really need to assume the “Markovian prior” ?
- Not necessarily ! We only need the ex-changeability (”conditional symmetry”).
- So even if we assume a-priori that the process is i.i.d, by collecting pair empirical distributions we can still argue that the conditioned process is Markov chain.
- Can continue this process : triple empirical distributions, ...
- Data-informed probability theory !

More general thoughts...

- The whole rationale of argument can be viewed from a measure-theoretic point of view.
- The initial **P** needs not be realistic for a particular system.
- The conditional symmetry that we revealed here is simply a result of the product structure of the underlying probability measure defining the process.
- Instead of the "hard" counting argument, can we obtain a rigorous "soft" perturbative argument for the 2nd level of combinatorial symmetry ?
- If so, we may extend our idea to the continuous-path processes.
- Obstacle : Cannot define a uniform measure on the space of continuous trajectories. Gaussian measure ? Wiener's construction ? Discard Probability Measure ?

Thank you for your attention !