Hypothesis testing for Panels of Semi-Markov Processes with parametric sojourn time distributions

Hervé Cardot, Cindy Frascolla

Hypothesis testing for Panels of Semi-Markov Processes with parametric sojourn time distributions

Hervé Cardot and Cindy Frascolla
Institut de Mathématiques de Bourgogne, UMR CNRS 5584,
Université de Bourgogne Franche-Comté

November 21, 2022

Abstract

This work deals with the asymptotic properties of maximum likelihood estimators for semi-Markov processes with parametric sojourn time distributions. It is motivated by the comparison, via a two-sample test procedure, of the distribution of two panels of qualitative trajectories modeled by semi-Markov processes and observed over a random number of transitions. Considering first one panel of growing size, we derive, under classical conditions, the convergence in probability of the estimators of the transition probabilities and the parameters of the sojourn time distributions as well as their asymptotic normality. We then consider panels of semi-Markov processes drawn from two different populations and study two-sample tests based on likelihood ratio. We also introduce a two-sample Wald type test. The finite sample performances of the proposed two-sample tests are evaluated with a brief simulation study.

Keywords: Likelihood ratio test; Panels of qualitative trajectories; Two-sample tests; Wald type test.

1 Introduction

This work is motivated by statistical experiments in sensory analysis in which the evolution of sensations over time can be modelled as trajectories of a qualitative stochastic process (see Lecuelle et al. (2018) for a detailed presentation of the experimental framework, see also Cardot et al. (2019)). In that context, Semi-Markov processes are shown to be relevant and parsimonious models, when considering a parametric characterization for the distribution of the sojourn times, to fit such kind of data. More generally, semi-Markov chains and semi-Markov processes, which allow to consider more flexible models for the distribution of the sojourn times than Markov chains, can provide interesting ways of modelling qualitative trajectories in many fields of science (see Barbu and Limnios (2008) and Limnios and Oprişan (2001) for general references). Estimators are based on the maximum likelihood

principle and our aim in this work is to derive the asymptotic properties of such estimators, with a particular focus on two-sample tests which are of great interest in food science and sensory analysis to compare two different products or two panels of subjects on the same product. Note that in that sensory analysis context, there is no censoring but the number of observed transitions can be considered as random and is not necessarily independent of the past of the trajectory.

A seminal work has been made by Billingsley (1961) to develop estimation procedures and to derive the asymptotic distribution of test statistics based on the likelihood ratio when the data are *i.i.d.* copies of a homogeneous Markov chain. However, considering Markov chains for qualitative trajectories imposes strong assumptions on the distribution of the sojourn times which may not be realistic in many applications, and particularly in sensory analysis. Asymptotic results for maximum likelihood estimators of semi-Markov processes are derived in Moore and Pyke (1968) based on long time behavior of a single trajectory and under conditions of irreducibility and recurrence. Considering two samples of hidden Markov models, Dannemann and Holzmann (2008) proved that, when the observation time tends to infinity, the asymptotic distribution of the likelihood ratio test statistic is a χ^2 law under the null hypothesis of equality of the distributions. As far as panels of semi-Markov chains are concerned, convergence results have been obtained for discrete time in Trevezas and Limnios (2011) in which the observation time is almost surely finite. In continuous time and under random censoring, Pons (2006) introduces non parametric approaches for the estimation of the distribution of the sojourn times and derives the asymptotic normality of the estimators. Almost sure consistency results are obtained in Barbu et al. (2017), for parametric sojourn time distributions which are closed under extrema (such as exponential, Weibull or Pareto distributions).

In this context of semi-Markov processes with parametric specifications for the distribution of the sojourn times, an empirical study has been performed in Frascolla et al. (2022) to evaluate the finite sample effectiveness of two samples testing procedures based on likelihood ratio statistics for sequences observed over random periods of time and drawn from two distinct populations. It has been noted that the likelihood ratio statistics is approximately distributed as a χ^2 law when the number of states is not too large and the sample size as well as the mean number of observed transitions are large enough.

The present work aims at deriving, in the same parametric framework, the asymptotic distribution of two-sample tests under the null hypothesis under general conditions on the sojourn time distributions. For that purpose, it is first needed to prove the asymptotic convergence of maximum likelihood estimators for the sojourn time distributions (Section 3). One novelty comes from the fact that there is a random number of observations used to build the estimators and the proofs are based on properties of the expected likelihood considering stopping times (see Gut (2009)) as well as more classical asymptotic tools (see

Newey and McFadden (1994) and Ferguson (1996)). Then, in Section 4, we consider two-sample tests of equality of distribution, and show under classic conditions that likelihood ratio statistics are asymptotically distributed as a χ^2 under the null hypothesis of equality, even if the observation protocols, described by stopping times, are not the same for the two panels. We also introduce a Wald type test of equality based on the asymptotic normality of the estimators and prove that under general conditions it is also asymptotically distributed as a χ^2 law under the null hypothesis of equality. Finally, a small simulation study is carried out in Section 5 to evaluate and compare these two approaches for finite samples. Concluding remarks are given in Section 6. All proofs are gathered in an Appendix.

2 Definitions, observed trajectories and likelihood

Notations are borrowed from Limnios and Oprişan (2001). We consider a stochastic process $Z = (Z_t)_{t\in\mathbb{R}_+}$ taking values in a finite state space $E = \{1, \ldots, D\}$ with $D < +\infty$. We denote by $J = (J_k)_{k\in\mathbb{N}}$ the successive visited states by Z and by $T = (T_k)_{k\in\mathbb{N}}$ the successive time points corresponding to a change of state. We also define $X = (X_k)_{k\in\mathbb{N}^*}$ with $X_k = T_k - T_{k-1}$ the successive sojourn times in the visited states. We assume that $Z = (Z_t)_{t\in\mathbb{R}_+}$ is a semi-Markov process (SMP) associated to (J, T), that is to say, $\forall j \in E$ and $t \in [0, +\infty)$,

$$\mathbb{P}(J_{k+1} = j, T_{k+1} - T_k \leq t \mid J_0, \ldots, J_k; T_0, \ldots, T_k) = \mathbb{P}(J_{k+1} = j, T_{k+1} - T_k \leq t \mid J_k) \quad (1)$$

We define $N(t) = \max\{k \in \mathbb{N} | T_k \leq t\}, t \in \mathbb{R}_+$ the counting process of the number of jumps in the time interval $(0, t]$. All along this work we assume that the SMP is regular, that is $\mathbb{P}(N(t) < \infty) = 1$ for all $t > 0$. We have $Z_t = J_{N(t)}$, for $t \geq 0$ and $J_k = Z_{T_k}$, for $k = 1, 2, \ldots$.

The law of the semi-Markov process Z is characterized by its initial distribution $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_D)$ with $\alpha_j = \mathbb{P}(J_0 = j)$, $j = 1, \ldots, D$ and its semi-Markov kernel,

$$Q_{ij}(t) = \mathbb{P}(J_k = j, X_k \leq t \mid J_{k-1} = i) \quad (2)$$

with the convention $X_0 = T_0 = 0$.

We denote by $\mathbf{P}$ the transition matrix of the embedded homogeneous Markov chain $(J_k)_{k\geq 1}$, with generic elements $p_{ij} = \mathbb{P}(J_k = j \mid J_{k-1} = i)$, for $i \neq j \in E \times E$. Note that by definition of the semi-Markov process, $p_{ii} = 0$, for all $i \in E$. We finally introduce, for $i \neq j \in E \times E$, the sojourn time cumulative distribution functions,

$$W_{ij}(t) = \mathbb{P}(X_k \leq t \mid J_{k-1} = i, J_k = j), \quad t \geq 0.$$

We have $Q_{ij}(t) = p_{ij}W_{ij}(t)$ as well as $\mathbb{P}(J_k = j) = \sum_{i\in E} \alpha_i p_{ij}^{(k)}$, where $p_{ij}^{(k)}$ is the generic element of matrix $\mathbf{P}^k$, that is to say, $p_{ij}^{(k)} = \left[\mathbf{P}^k\right]_{i,j}$, $(i, j) \in E \times E$. We adopt the convention that $\mathbf{P}^0$ is the identity matrix, so that $\mathbb{P}(J_0 = j) = \sum_{i\in E} \alpha_i p_{ij}^{(0)} = \alpha_j$.

We consider in this work parametric sojourn time distribution functions and we denote by $f(t, \theta_{ij})$ the densities (resp. the probabilities) if time is continuous (resp. if time is discrete), depending on the vector of parameters $\theta_{ij} \in \mathbb{R}^d$, with $i \neq j \in E \times E$. We define $\boldsymbol{\theta} = (\theta_{ij}, i \neq j \in E \times E)$ the set of parameters related to the sojourn time distributions.

The distribution of Z is thus characterized by the vector of parameters $(\boldsymbol{\alpha}_0, \mathbf{P}_0, \boldsymbol{\theta}_0)$. Taking account of the constraints that naturally arise, $\sum_{j=1}^{D} \alpha_j = 1$, $\sum_{j=1}^{D} p_{ij} = 1$ and $p_{ii} = 0$ for $i = 1, \ldots, D$, the total number of unknown parameters is equal to $D-1+D(D-2)+dD(D-1)$ when there is no absorbing state. If there is an absorbing state and if we suppose that $\alpha_D = 0$, the number of unknown parameters is equal to $D-2+(D-1)(D-2)+d(D-1)^2$ since $P_{DD} = 1$ and $P_{Dj} = 0$ for all $j \neq D$, with no associated sojourn time distribution.

2.1 Observation protocol and likelihood

With real experiments, we do not observe sequences having an infinite number of transitions, and, as in sensory analysis experiments, we can suppose that for each sequence the number of observed transitions is random and not necessarily independent to the past of the trajectory, but without any censoring for the sojourn times. The observation process is thus stopped after a random number of M transitions which occur almost surely in a finite time. Given that $M = m$, for a strictly positive integer m, we have access to the observation of $\boldsymbol{S} = \{j_0, x_1, j_1, \ldots, j_{m-1}, x_m, j_m\}$, whose likelihood is equal to

$$\mathcal{L}(\boldsymbol{S}, \boldsymbol{\alpha}, \mathbf{P}, \boldsymbol{\theta}) = \alpha_{j_0} \prod_{k=1}^{m} p_{j_{k-1}j_k} f\left(x_k, \theta_{j_{k-1}j_k}\right). \tag{3}$$

The integer valued random variable M can be supposed to be independent of the trajectory $(J_k, X_k)_{\{k \geq 0\}}$ or it can be supposed to be a stopping time with respect to the increasing sequence of sub-σ-algebra $\mathcal{F}_k = \sigma(J_0, X_0, \ldots, J_k, X_k)$. Realistic examples of stopping times, in our context are

- $M(t)$ is the number of visited states until time t, $M(t) = \inf\{k \in \mathbb{N} | \sum_{j=1}^{k} X_j \geq t\}$. It is well known for renewal processes that if $\mathbb{E}(X) > 0$ then $M(t)$ is finite almost surely. We also suppose that $\mathbb{P}[M(t) \geq 1] > 0$ so that at least one transition can be observed with non null probability.

- M_D is the number of visited states before absorption, defined as follows $M_D = \inf\{k \in \mathbb{N} | J_k = D\}$, where we define the last state $\{D\}$ to be the absorbing state. In case, given that $M = \tau$, the likelihood of a trajectory $\boldsymbol{S} = \{j_0, x_1, j_1, \ldots, j_{\tau-1}, x_\tau, D\}$ is equal to

$$\mathcal{L}(\boldsymbol{S}, \boldsymbol{\alpha}, \mathbf{P}, \boldsymbol{\theta}) = \alpha_{j_0} \left[\prod_{k=1}^{\tau-1} p_{j_{k-1}j_k} f(x_k, \theta_{j_{k-1}j_k}) \right] p_{j_{\tau-1}D} f(x_\tau, \theta_{j_{\tau-1}D}). \tag{4}$$

For Markov chains having one absorbing state (see Kemeny and Snell (1976), Chapter 3 for details), the transition matrix $\mathbf{P}$ can be decomposed as follows

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & 1 \end{pmatrix} \tag{5}$$

where $\mathbf{Q}$ is the $(D-1) \times (D-1)$ matrix giving the transition probabilities among the no absorbing states, and $\mathbf{R}$ is the vector with generic elements $p_{0_i D}$, $i \in \{1, \ldots, D-1\}$.

The matrix $\mathbf{I} - \mathbf{Q}$, with $\mathbf{I}$ the identity matrix, has an inverse (see Theorem 3.2.1 in Kemeny and Snell (1976)), denoted by $\mathbf{F} = (\mathbf{I} - \mathbf{Q})^{-1}$ and called the fundamental matrix. The generic element $[\mathbf{F}]_{ij}$ is the expected number of times the chain is in state j given that it started in state i. We can deduce with Theorem 3.3.5 in Kemeny and Snell (1976) that $\mathbb{E}[M_D | J_0 = i] = \sum_{j=1}^{D-1} F_{ij} < +\infty$ and $\mathbb{E}[M_D] = \sum_{i=1}^{D-1} \mathbb{E}[M_D | J_0 = i] \, \alpha_{0_i}$.

We suppose now that we have a panel of n independent trajectories of Z, denoted by $\boldsymbol{S_1}, \ldots, \boldsymbol{S_n}$. We define $\hat{Q}(\boldsymbol{\alpha}, \mathbf{P}, \boldsymbol{\theta})$ the average value, over the trajectories $\boldsymbol{S_1}, \ldots, \boldsymbol{S_n}$, of the log-likelihood,

$$\widehat{Q}(\boldsymbol{\alpha}, \mathbf{P}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{\ell=1}^{n} \ln \mathcal{L}(\boldsymbol{S_\ell}; \boldsymbol{\alpha}, \mathbf{P}, \boldsymbol{\theta}). \tag{6}$$

When it exists, a maximum likelihood estimator of $(\boldsymbol{\alpha}, \mathbf{P}, \boldsymbol{\theta})$ is denoted by $(\widehat{\boldsymbol{\alpha}}, \widehat{\mathbf{P}}, \widehat{\boldsymbol{\theta}})$ and satisfies $\widehat{Q}(\widehat{\boldsymbol{\alpha}}, \widehat{\mathbf{P}}, \widehat{\boldsymbol{\theta}}) \geq \widehat{Q}(\boldsymbol{\alpha}, \mathbf{P}, \boldsymbol{\theta})$, for all possible set parameters $(\boldsymbol{\alpha}, \mathbf{P}, \boldsymbol{\theta})$.

We can remark that taking account of the particular multiplicative form of (3), criterion (6) can be decomposed into three parts,

$$\widehat{Q}(\boldsymbol{\alpha}, \mathbf{P}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{\ell=1}^{n} \ln \left(\alpha_{j_0^{(\ell)}} \right) + \widehat{Q}_{\mathbf{P}}(\mathbf{P}) + \widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta}). \tag{7}$$

A direct consequence is that the maximum likelihood estimators of $\mathbf{P}$ and $\boldsymbol{\theta}$ can be computed independently by looking separately for the maximum of $\widehat{Q}_{\mathbf{P}}(\mathbf{P})$ and $\widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$.

2.2 Maximum likelihood estimators

The maximum likelihood estimator $\widehat{\boldsymbol{\alpha}} = (\widehat{\alpha}_1, \ldots, \widehat{\alpha}_D)$ of the vector of initialization probabilities $\boldsymbol{\alpha}_0 = (\alpha_{0_1}, \ldots, \alpha_{0_D})$ is, in all the considered cases, given by

$$\widehat{\alpha}_j = \frac{1}{n} \sum_{\ell=1}^{n} \mathbf{1}_{\{J_0^{(\ell)} = j\}}, \quad j = 1, \ldots, D \tag{8}$$

where $\mathbf{1}_{\{J_0^{(\ell)} = i\}}$ is the indicator function taking value 1 if the first visited state by sequence $\mathbf{S}_\ell$ is $\{i\}$ and zero else. The estimator $\widehat{\boldsymbol{\alpha}}$ is simply the maximum likelihood estimator of the vector of parameters $\boldsymbol{\alpha}_0$ of a multinomial distribution and it is shown with classical arguments (see *e.g.* Anderson and Goodman (1957) or Agresti (2002)) that, under assumption $\mathbf{A}_1$ given below, $\widehat{\boldsymbol{\alpha}}$ is a consistent estimator of $\boldsymbol{\alpha}_0$ when n tends to infinity

and that $\sqrt{n}\,(\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)$ converges in distribution to a centered Gaussian law with (singular) covariance matrix $\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}$, whose generic elements are $[\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}]_{ij} = -\alpha_{0_i}\alpha_{0_j}$ if $i \neq j$ and $[\boldsymbol{\Gamma}_{\boldsymbol{\alpha}}]_{ii} = \alpha_{0_i}(1 - \alpha_{0_i})$.

From now on, we focus on the estimation of the parameters related to the transition probabilities $\mathbf{P}_0$ and the set of parameters $\boldsymbol{\theta}_0$ characterizing the laws of the sojourn times. For $i \in E$, we denote by $\mathbf{p}_i = (p_{ij}, j \neq i) \in \mathbb{R}^{(D-1)}$ the vector of transition probabilities from state $\{i\}$ to the other states and by $\widehat{\mathbf{p}}_i = (\widehat{p}_{i1}, \ldots, \widehat{p}_{iD})$ its maximum likelihood estimator. We denote by $\mathbf{p}_0 = (\mathbf{p}_1, \cdots, \mathbf{p}_D) \in \mathbb{R}^{D(D-1)}$ the vector obtained by the concatenation of $\mathbf{p}_1, \cdots, \mathbf{p}_D$ and by $\widehat{\mathbf{p}}$ its maximum likelihood estimator. In presence of an absorbing state $\{D\}$, we only consider $\mathbf{p}_0 = (\mathbf{p}_1, \ldots, \mathbf{p}_{D-1})$ since $\mathbf{p}_D = (0, \ldots, 0, 1)$.

The elements of $\boldsymbol{\theta}_0$ are rearranged so that they form a $D(D-1)d$ dimensional vector, with $\boldsymbol{\theta}_0 = (\theta_{1,2}, \ldots, \theta_{1,D}, \ldots, \theta_{D,D-1})$, and we denote by $\widehat{\boldsymbol{\theta}}$ its maximum likelihood estimator. In presence of an absorbing state, we only consider the $(D-1)(D-1)d$ vector $\boldsymbol{\theta}_0 = (\theta_{1,2}, \ldots, \theta_{1,D}, \ldots, \theta_{D-1,D})$.

For each trajectory $\boldsymbol{S_\ell}$, we denote by m_ℓ the number of observed transitions. We define

$$N_i^{(\ell)} = \sum_{k=0}^{m_\ell - 1} \mathbf{1}_{\{J_k^{(\ell)} = i\}} \tag{9}$$

the number of times state $\{i\}$ is reached in sequence $\boldsymbol{S_\ell}$ and by

$$N_{ij}^{(\ell)} = \sum_{k=0}^{m_\ell - 1} \mathbf{1}_{\{J_k^{(\ell)} = i, J_{k+1}^{(\ell)} = j\}} \tag{10}$$

the number of observed transitions from $\{i\}$ to $\{j\}$. When $N_{ij}^{(\ell)} \geq 1$, we denote for $k = 1, \ldots, N_{ij}^{(\ell)}$, by $x_{ij}^{(\ell,k)}$ the sojourn time, at state $\{i\}$ and before moving to state $\{j\}$, during the kth visit of sequence S_ℓ. We have

$$\widehat{Q}_{\mathbf{P}}(\mathbf{P}) = \frac{1}{n} \sum_{\ell=1}^{n} \left\{ \sum_{i \in E} \sum_{\substack{j=1 \\ j \neq i}}^{D} \left[N_{ij}^{(\ell)} \ln(p_{ij}) \right] \right\} \tag{11}$$

and

$$\widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{\ell=1}^{n} \sum_{\substack{i,j \in E \\ j \neq i}} \sum_{k=1}^{N_{ij}^{(\ell)}} \ln\left(f(x_{ij}^{(\ell,k)}, \theta_{ij}) \right) \tag{12}$$

with the convention that $\sum_{k=1}^{N_{ij}^{(\ell)}} \ln\left(f(x_{ij}^{(\ell,k)}, \theta_{ij}) \right) = 0$ when $N_{ij}^{(\ell)} = 0$.

The maximum likelihood estimators of the elements $\widehat{p}_{ij}$ of the matrix of transition probabilities $\mathbf{P}$ are obtained by solving the constrained optimization problem,

$$\begin{aligned} &\max_{\mathbf{P}} \widehat{Q}_{\mathbf{P}}(\mathbf{P}) \\ &\text{subject to} \sum_{j \in E} p_{ij} = 1, \quad \forall\, i \in E. \end{aligned}$$

Introducing Lagrange multipliers, we consider the Lagrangian,

$$\widehat{Q}_{\mathbf{P}}(\mathbf{P}, \lambda_1, \ldots, \lambda_D) = \frac{1}{n} \sum_{\ell=1}^{n} \left\{ \sum_{i \in E} \sum_{\substack{j=1 \\ j \neq i}}^{D} \left[N_{ij}^{(\ell)} \ln(p_{ij}) \right] \right\} + \sum_{i \in E} \lambda_i \left(\sum_{j \in E} p_{ij} - 1 \right). \tag{13}$$

The maximum likelihood estimators of the elements $\widehat{p}_{ij}$ of the matrix of transition probabilities $\mathbf{P}$ are obtained by finding the roots of the gradient $\nabla_{\mathbf{P}} \widehat{Q}_{\mathbf{P}}(\mathbf{P}, \lambda_1, \ldots, \lambda_D)$ whose elements are for $j \neq D$ and $j \neq i$,

$$\frac{\partial \widehat{Q}_{\mathbf{P}}(\mathbf{P}, \lambda_1, \ldots, \lambda_D)}{\partial p_{ij}} = \frac{1}{n} \sum_{\ell=1}^{n} \left(\frac{N_{ij}^{(\ell)}}{p_{ij}} \right) + \lambda_i \tag{14}$$

$$\frac{\partial \widehat{Q}_{\mathbf{P}}(\mathbf{P}, \lambda_1, \ldots, \lambda_D)}{\partial \lambda_i} = \sum_{j \in E} p_{ij} - 1. \tag{15}$$

The solutions are (see Anderson and Goodman (1957) or Billingsley (1961))

$$\widehat{p}_{ij} = \frac{\sum_{\ell=1}^{n} N_{ij}^{(\ell)}}{\sum_{\ell=1}^{n} N_{i}^{(\ell)}}, \quad \text{if } \sum_{\ell=1}^{n} N_{i}^{(\ell)} \geq 1 \tag{16}$$

setting $\widehat{p}_{ij} = 0$ if $\sum_{\ell=1}^{n} N_{i}^{(\ell)} = 0$.

Maximum likelihood estimators $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ are obtained by setting to zero the gradient $\nabla_{\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}})$ of $\widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$, that is to say

$$\begin{aligned} \nabla_{\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) &= \frac{1}{n} \sum_{\ell=1}^{n} \sum_{\substack{i,j \in E \\ j \neq i}} \sum_{k=1}^{N_{ij}^{(\ell)}} \nabla_{\theta} \ln \left(f(x_{ij}^{(\ell,k)}, \widehat{\theta}_{ij}) \right) \\ &= 0. \end{aligned} \tag{17}$$

There is in general no explicit solution to such system of equations and a solution $\widehat{\boldsymbol{\theta}}$ to this implicit equation is obtained by numerical iterative techniques.

3 Assumptions and convergence properties

The convergence results are based on properties of semi-Markov processes and classical conditions on the distribution of the sojourn times that ensure uniqueness and consistency of maximum likelihood estimators (see *e.g* Newey and McFadden (1994)) as well as the Wald identity (recalled in Theorem 6.1 in the Appendix, see for example Theorem 5.3, Chapter 1 in Gut (2009)) to deal with random sample sizes and stopping times. We denote by $\|.\|$ the Euclidean norm for vectors in $\mathbb{R}^d$ as well as the spectral norm for real matrices. For a real valued function $h(x, \theta)$ depending on $x \in \mathbb{R}$ and a parameter $\theta \in \mathbb{R}^d$, we denote its gradient by $\nabla_{\theta} h(x, \theta)$ and by $\nabla_{\theta\theta} h(x, \theta)$ the corresponding Hessian matrix evaluated at θ. We suppose that

$\mathbf{A}_1$: $\alpha_{0_j} > 0, \quad \forall j \in E.$

$\mathbf{A}_2$: $p_{0_{ij}} > 0, \quad \forall (i, j \neq i) \in E \times E.$

$\mathbf{A}_3$: $\theta_{0_{ij}} \in \mathbf{\Theta}$, $\forall (i, j \neq i) \in E \times E$, where $\mathbf{\Theta}$ is a compact set in $\mathbb{R}^d$.

$\mathbf{A}_4$: $f(., \theta) \neq f(., \theta_0)$ as far as $\theta \neq \theta_0$, with $(\theta, \theta_0) \in \mathbf{\Theta} \times \mathbf{\Theta}$.

$\mathbf{A}_5$: $\ln f(X, \theta)$ is continuous at each $\theta \in \mathbf{\Theta}$ with probability one, and $\mathbb{E}_{\boldsymbol{\theta}_0} \left[\sup_{\theta \in \mathbf{\Theta}} |\ln f(X, \theta)|\right] < +\infty$.

$\mathbf{A}_6$: We assume that $\forall (i, j \neq i) \in E \times E$,

- i $\theta_{0_{ij}}$ belongs to the interior of $\mathbf{\Theta}$
- ii $f(X, \theta)$ is twice continuously differentiable and $f(X, \theta) > 0$ in a neighborhood $\mathcal{N}_{ij}$ of $\theta_{0_{ij}}$
- iii $\int \sup_{\theta \in \mathcal{N}_{ij}} \|\nabla_\theta f(x, \theta)\| \, dx < +\infty$
- iv $\int \sup_{\theta \in \mathcal{N}_{ij}} \|\nabla_{\theta\theta} f(x, \theta)\| \, dx < +\infty$
- v $\mathbb{E}_{\boldsymbol{\theta}_0} \left[\nabla_\theta \ln f(X, \theta) \left(\nabla_\theta \ln f(X, \theta)\right)^\top\right]$ is a non singular matrix
- vi $\mathbb{E}_{\boldsymbol{\theta}_0} \left[\sup_{\theta \in \mathcal{N}_{ij}} \|\nabla_{\theta\theta} \ln f(X, \theta)\|\right] < +\infty$.

Assumption $\mathbf{A}_1$, which also appears in Trevezas and Limnios (2011), simply ensures that every state can be reached during the first jump and simplifies the presentation of the results. Since $\sum_{j=1}^{D} \alpha_{0_j} = 1$, it also implies that $\alpha_{0_j} < 1$, $\forall j \in E$. Assumption $\mathbf{A}_2$ ensures that every state is accessible in one transition. It allows simpler calculations and could be weakened at the expense of heavier notations. Assumption $\mathbf{A}_3$ is an usual assumption of compactness for maximum likelihood estimators. It could be weakened under additional concavity conditions of the likelihood (see Hjort and Pollard (2011)). Hypothesis $\mathbf{A}_4$ is an identification condition which ensures uniqueness of the maximum of the expected likelihood at the true value of the parameter. Conditions appearing in $\mathbf{A}_5$ allow to obtain a uniform result of convergence of the log likelihood function and to deduce almost sure convergence of the maximum likelihood estimators for the parameters related to the sojourn time distributions. The set of conditions appearing in $\mathbf{A}_6$ are classical for getting the asymptotic normality of maximum likelihood estimators in an *i.i.d* context and are similar to those given in Theorem 3.3 in Newey and McFadden (1994). Note that, to be slightly more general, the conditions $\mathbf{A}_5$ and $\mathbf{A}_6$ could certainly be replaced by a quadratic mean differentiability assumption (see *e.g.* Chapter 5 in van der Vaart (1998) or Chapter 12 in Lehmann and Romano (2005)) on the law of the sojourn times.

Note that in presence of an absorbing state $\{D\}$, assumptions $\mathbf{A}_1$ and $\mathbf{A}_2$ only deal with the unknown transition probabilities and should be understood as follows $\mathbf{A}_1$: $\alpha_{0_j} >$

$0, \quad \forall j \in E\backslash\{D\}$ and $\mathbf{A}_2$: $p_{0_{ij}} > 0, \quad \forall (i, j \neq i) \in E\backslash\{D\} \times E$. The same restriction holds on the indices (i, j) for assumptions $\mathbf{A}_3$ and $\mathbf{A}_6$.

Theorem 3.1. *If hypotheses* $\mathbf{A}_1$ *to* $\mathbf{A}_5$ *hold and if individual sequences are observed during* M *transitions with* $\mathbb{E}[M] < +\infty$*, then, as* $n \to \infty$*,*

$$\widehat{\mathbf{P}} \to \mathbf{P}_0 \quad \textit{almost surely,}$$
$$\widehat{\boldsymbol{\theta}} \to \boldsymbol{\theta}_0 \quad \textit{almost surely.}$$

As far as the asymptotic distribution of $\widehat{\boldsymbol{\theta}}$ is concerned, the additional assumptions given in $\mathbf{A}_6$ are classical assumptions in maximum likelihood theory which ensure asymptotic normality of the estimators of each θ_{ij} based on independent copies of X_{ij} (see for example Theorem 3.3 in Newey and McFadden (1994)).

Theorem 3.2. *If hypotheses* $\mathbf{A}_1$ *to* $\mathbf{A}_6$ *hold and if* $\mathbb{E}[M] < +\infty$*, then as* n *tends to infinity,*

$$\sqrt{n}\left(\begin{pmatrix}\widehat{\mathbf{p}}\\ \widehat{\boldsymbol{\theta}}\end{pmatrix} - \begin{pmatrix}\mathbf{p}_0\\ \boldsymbol{\theta}_0\end{pmatrix}\right) \rightsquigarrow \mathcal{N}\left(0, \boldsymbol{\Gamma}^M_{p_0,\theta_0}\right)$$

with $\boldsymbol{\Gamma}^M_{p_0,\theta_0} = \begin{pmatrix}\boldsymbol{\Gamma}^M_{p_0} & 0\\ 0 & \boldsymbol{\Gamma}^M_{\theta_0}\end{pmatrix}$ *and* $\boldsymbol{\Gamma}^M_{p_0}$ *and* $\boldsymbol{\Gamma}^M_{\theta_0}$ *are block diagonal matrices. If* $M = M(t)$ *then*

$$\boldsymbol{\Gamma}^M_{p_0} = \begin{pmatrix} \frac{1}{\mathbb{E}[N_1^M]}\boldsymbol{\Gamma}_{p_1} & 0 & \cdots & 0\\ 0 & \ddots & 0 & \vdots\\ \vdots & 0 & \ddots & 0\\ 0 & \cdots & 0 & \frac{1}{\mathbb{E}[N_D^M]}\boldsymbol{\Gamma}_{p_D}\end{pmatrix}$$

with (singular) matrices $\boldsymbol{\Gamma}_{p_i}$*, for* $i = 1, \ldots, D$*, having diagonal elements* $[\boldsymbol{\Gamma}_{p_i}]_{jj} = p_{0ij}(1 - p_{0ij})$ *if* $j \neq i$ *and* $[\boldsymbol{\Gamma}_{p_i}]_{kj} = -p_{0ij}p_{0ik}$ *if* $j \neq k$*. Matrix* $\boldsymbol{\Gamma}^M_{\theta_0}$ *is also block diagonal, with matrices on the diagonal*

$$\boldsymbol{\Gamma}^M_{\theta_{ij}} = \frac{1}{p_{0_{ij}}\mathbb{E}\left[N_i^M\right]}\left(-\mathbb{E}\left[\nabla_{\theta_{ij}\theta_{ij}} \ln\left(f(X_{ij}; \theta_{0_{ij}})\right)\right]\right)^{-1}.$$

If $M = M_D$ *then* $\boldsymbol{\Gamma}^M_{p_0}$ *and* $\boldsymbol{\Gamma}^M_{\theta_0}$ *have the same expression but are made of* $D - 1$ *matrices on the diagonal, excluding the part corresponding to the absorbing state.*

The proofs of these two theorems are given in the Appendix. Note that the asymptotic distribution of $\widehat{\mathbf{p}}$ has already been derived in the simpler design in which we observe exactly m transitions for each trajectory in Anderson and Goodman (1957) and for sequences $\boldsymbol{S}$ censored at time t in Trevezas and Limnios (2011).

Remark. As an anonymous referee pointed out, it could be possible to consider that we observe a deterministic number M_n of transitions for each trajectory, with M_n tending to infinity as n tends to infinity. A closer look at Theorem 3.2 shows that the asymptotic covariance matrix $\mathbf{\Gamma}_{p_0}^M$ is a block diagonal matrix made of D covariance matrices which are "normalized", for $j = 1, \ldots, D$, by $1/\mathbb{E}[N_j^M]$, the inverse of the expected number of times a trajectory reaches state j. The same normalizing factors appear in $\mathbf{\Gamma}_{\theta_0}^M$. We have

$$\begin{aligned}\mathbb{E}[N_j^{M_n}] &= \sum_{k=0}^{M_n-1} \mathbb{P}[J_k = j] \\ &= \sum_{i \in E} \sum_{k=0}^{M_n-1} \alpha_i p_{ij}^{(k)}.\end{aligned}$$

Under assumption $\mathbf{A}_2$, we have that the embedded homogeneous Markov chain $(J_k)_{k\geq 0}$ is an irreducible positive recurrent Markov chain, which consequently admits an invariant probability measure $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_D)$ satisfying

$$\lim_{k\to\infty} p_{ij}^{(k)} = \pi_j.$$

By Cesaro's mean theorem, and the fact that $\sum_{i\in E} \alpha_i = 1$, we deduce that

$$\lim_{n\to\infty} \frac{1}{M_n} \mathbb{E}[N_j^{M_n}] = \pi_j$$

so that, when M_n is large, $\mathbb{E}[N_j^{M_n}] \approx \pi_j M_n$. This means that, if $(M_n)_{n\geq 1}$ is a sequence of integers tending to infinity, the $\sqrt{n}$ normalizing factor in Theorem 3.2 should be replaced by $\sqrt{nM_n}$ and $1/\mathbb{E}[N_j^M]$ should be replaced by $1/\pi_j$.

4 Likelihood ratio and Wald type tests

Different strategies, which are more or less asymptotically equivalent, can be considered for hypothesis testing in this maximum likelihood parametric context, as explained in Newey and McFadden (1994). One advantage of the approach based on the likelihood ratio is that it does not require to have at hand an explicit estimate of the variance of the estimators, which may be a not so simple task in our setting, and it generally leads to a simple asymptotic chi-square distribution under general conditions. A simple introduction to likelihood ratio test is given in Ferguson (1996), Chapter 22.

Testing the equality of the initialization probabilities, that is to say the null hypothesis $H_0 : \boldsymbol{\alpha}^1 = \boldsymbol{\alpha}^2$, with a likelihood ratio test approach follows directly from standard tools which do not involve the dynamics of the trajectories and is thus not presented here (see example 16.1 in van der Vaart (1998)). We focus in the following on testing hypotheses on the transition probabilities and sojourn time distributions. Define the set R as the set of all possible transition matrices satisfying assumption $\mathbf{A}_2$. Consider R_0 a subset of R and

Θ_0 a subset of Θ and suppose now we would like to test $H_0 : (\mathbf{p}, \boldsymbol{\theta}) \in R_0 \times \Theta_0$ against the alternative $H_1 : (\mathbf{p}, \boldsymbol{\theta}) \notin R_0 \times \Theta_0$.

We consider the likelihood ratio statistics defined as follows

$$\lambda_n = \frac{\sup_{(\mathbf{p},\boldsymbol{\theta})\in R_0\times\Theta_0} \prod_{\ell=1}^n \mathcal{L}(\boldsymbol{S}_\ell, \boldsymbol{\alpha}, \mathbf{p}, \boldsymbol{\theta})}{\sup_{(\mathbf{p},\boldsymbol{\theta})\in R\times\Theta} \prod_{\ell=1}^n \mathcal{L}(\boldsymbol{S}_\ell, \boldsymbol{\alpha}, \mathbf{p}, \boldsymbol{\theta})}. \tag{18}$$

Note that λ_n does not depend on $\boldsymbol{\alpha}$ because of the multiplicative structure of the log-likelihood $\mathcal{L}$ (see equation 7).

4.1 Simple hypothesis testing

We want to test the simple null hypothesis $H_0 : \ (\boldsymbol{p}, \boldsymbol{\theta}) = (\boldsymbol{p}_0, \boldsymbol{\theta}_0)$, for some values $\boldsymbol{p}_0$ for $\mathbf{p}$ and $\boldsymbol{\theta}_0$ for $\boldsymbol{\theta}$ given in advance. We consider the following likelihood ratio,

$$\lambda_n = \frac{\prod_{\ell=1}^n \mathcal{L}(\boldsymbol{S}_\ell, \boldsymbol{p}_0, \boldsymbol{\theta_0})}{\prod_{\ell=1}^n \mathcal{L}(\boldsymbol{S}_\ell, \widehat{\mathbf{p}}, \widehat{\boldsymbol{\theta}})}, \tag{19}$$

whose asymptotic distribution is given, when H_0 is true, in the theorem below.

Theorem 4.1. *Suppose that the conditions of Theorem 3.2 are in force, if the null hypothesis $H_0 : \ (\boldsymbol{p}, \boldsymbol{\theta}) = (\boldsymbol{p}_0, \boldsymbol{\theta}_0)$ is true then, as n tends to infinity,*

$$-2\ln\lambda_n \rightsquigarrow \chi^2_{DoF}$$

where the degrees of freedom are equal to $DoF = D(D-2) + D(D-1)d$ when there is no absorbing state and $DoF = (D-1)(D-2) + (D-1)^2 d$ when there is an absorbing state.

Note that partial tests $H_0^\theta : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ or $H_0^p : \ \boldsymbol{p} = \boldsymbol{p}_0$ dealing only with the transition probabilities or the sojourn time distributions can be performed easily. It can be directly deduced, thanks to the particular structure (7) of the log likelihood, that the test statistics $-2\ln\lambda_n$ is asymptotically distributed as a χ^2 with a degree of freedom equal to the number of constraints imposed under the null hypothesis. If we consider the null hypothesis $H_0^\theta : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, then $-2\ln\lambda_n \rightsquigarrow \chi^2_{D(D-1)d}$ when there is no absorbing state and $-2\ln\lambda_n \rightsquigarrow \chi^2_{(D-1)^2 d}$ when the trajectories are observed until absorption. If we consider the null hypothesis $H_0^p : \boldsymbol{p} = \boldsymbol{p}_0$, then $-2\ln\lambda_n \rightsquigarrow \chi^2_{D(D-2)}$ when there is no absorbing state and $-2\ln\lambda_n \rightsquigarrow \chi^2_{(D-1)(D-2)}$ in presence of an absorbing state.

4.2 Two-sample tests

Suppose now we have two samples of respectively n_1 and n_2 trajectories, $(\boldsymbol{S}^1_\ell)_{\ell=1,\dots,n_1}$ and $(\boldsymbol{S}^2_\ell)_{\ell=1,\dots,n_2}$ drawn from two distinct populations whose probability distributions are characterized by the parameters $(\boldsymbol{\alpha}^1, \mathbf{p}^1, \boldsymbol{\theta}^1)$ and $(\boldsymbol{\alpha}^2, \mathbf{p}^2, \boldsymbol{\theta}^2)$. We would like to test the equality of the distributions, that is to say test the null hypothesis

$$H_0 : (\mathbf{p}^1, \boldsymbol{\theta}^1) = (\mathbf{p}^2, \boldsymbol{\theta}^2). \tag{20}$$

4.2.1 A two-sample Wald-type test

A way of testing the equality of the distribution of two semi-Markov processes consists in considering Wald testing approaches based on the asymptotic distribution of the estimators under the null hypothesis. We first state the following Lemma.

Lemma 4.1. *Suppose that the conditions of Theorem 3.2 are in force and* $\frac{n_2}{n_1+n_2} \to f \in (0,1)$ *as* n_1, n_2 *tend to infinity,*

$$\sqrt{\frac{n_1 n_2}{n_1+n_2}} \begin{pmatrix} \widehat{\boldsymbol{p}}^{(1)} - \widehat{\boldsymbol{p}}^{(2)} - (\mathbf{p}^1 - \mathbf{p}^2) \\ \widehat{\boldsymbol{\theta}}^{(1)} - \widehat{\boldsymbol{\theta}}^{(2)} - (\boldsymbol{\theta}^1 - \boldsymbol{\theta}^2) \end{pmatrix} \rightsquigarrow \mathcal{N}\left(0, \mathbf{\Gamma}^f_{p,\theta}\right)$$

where $\mathbf{\Gamma}^f_{p,\theta} = \begin{pmatrix} \mathbf{\Gamma}^f_p & 0 \\ 0 & \mathbf{\Gamma}^f_\theta \end{pmatrix}$, *with* $\mathbf{\Gamma}^f_p = f\mathbf{\Gamma}^{M_1}_{p^1} + (1-f)\mathbf{\Gamma}^{M_2}_{p^2}$ *and* $\mathbf{\Gamma}^f_\theta = f\mathbf{\Gamma}^{M_1}_{\theta^1} + (1-f)\mathbf{\Gamma}^{M_2}_{\theta^2}$.

Based on Lemma 4.1, we are now able to build a Wald-type test statistic to test the hypothesis of equality (20). Consider

$$W^{p,\theta}_{n_1,n_2} = \frac{n_1 n_2}{n_1+n_2} \begin{pmatrix} \widehat{\boldsymbol{p}}^{(1)} - \widehat{\boldsymbol{p}}^{(2)} \\ \widehat{\boldsymbol{\theta}}^{(1)} - \widehat{\boldsymbol{\theta}}^{(2)} \end{pmatrix}^\top \left(\widehat{\mathbf{\Gamma}}^{f_n}_{p,\theta}\right)^{-1} \begin{pmatrix} \widehat{\boldsymbol{p}}^{(1)} - \widehat{\boldsymbol{p}}^{(2)} \\ \widehat{\boldsymbol{\theta}}^{(1)} - \widehat{\boldsymbol{\theta}}^{(2)} \end{pmatrix} \tag{21}$$

where

$$\widehat{\mathbf{\Gamma}}^{f_n}_{p,\theta} = \frac{n_2}{n_1+n_2} \begin{pmatrix} \widehat{\mathbf{\Gamma}}^{M_1}_p & 0 \\ 0 & \widehat{\mathbf{\Gamma}}^{M_1}_\theta \end{pmatrix} + \frac{n_1}{n_1+n_2} \begin{pmatrix} \widehat{\mathbf{\Gamma}}^{M_2}_p & 0 \\ 0 & \widehat{\mathbf{\Gamma}}^{M_2}_\theta \end{pmatrix} \tag{22}$$

and the estimators $\widehat{\mathbf{\Gamma}}^{M_1}_p$ and $\widehat{\mathbf{\Gamma}}^{M_2}_p$ (resp. $\widehat{\mathbf{\Gamma}}^{M_1}_\theta$ and $\widehat{\mathbf{\Gamma}}^{M_2}_\theta$) are obtained by replacing in the expression of the asymptotic covariance matrices the unknown values θ_{ij} and p_{ij} by their maximum likelihood estimators $\widehat{\theta}_{ij}$ and $\widehat{p}_{ij}$ computed under the null hypothesis and the expected number of times state $\{i\}$ is reached, $\mathbb{E}[N_i^{M_1}]$ (resp. $\mathbb{E}[N_i^{M_2}]$) by its empirical counterpart $n_1^{-1}\sum_{\ell=1}^{n_1} N_i^{1,\ell}$ (resp. $n_2^{-1}\sum_{\ell=1}^{n_2} N_i^{2,\ell}$). We recall that matrix $\widehat{\mathbf{\Gamma}}^M_p$ is singular, thus we consider its generalized inverse in the computation of $W^{p,\theta}_{n_1,n_2}$.

We can now state the following theorem.

Theorem 4.2. *Suppose that the conditions of Theorem 3.2 are in force and* $\frac{n_2}{n_1+n_2} \to f \in (0,1)$, *as* n_1 *and* n_2 *tend to infinity. If the null hypothesis* $H_0 :\ (\mathbf{p}^1, \boldsymbol{\theta}^1) = (\mathbf{p}^2, \boldsymbol{\theta}^2)$ *is true then*

$$W^{p,\theta}_{n_1,n_2} \rightsquigarrow \chi^2_{DoF}$$

where the degrees of freedom are equal to $DoF = D(D-2) + D(D-1)d$ *when there is no absorbing state and* $DoF = (D-1)(D-2) + (D-1)^2 d$ *when there is an absorbing state.*

Note that this Wald type test strategy can also be easily employed for testing partial equality hypotheses, dealing only with a subset of the parameters. The most natural partial hypotheses that we can consider are $H_0^p:\ \mathbf{p}_1 = \mathbf{p}_2$ or $H_0^\theta:\ \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. In that case, we can introduce the following test statistics

$$W^p_{n_1,n_2} = \frac{n_1 n_2}{n_1+n_2}\left(\widehat{\boldsymbol{p}}^{(1)} - \widehat{\boldsymbol{p}}^{(2)}\right)^\top \left(\frac{n_2}{n_1+n_2}\widehat{\boldsymbol{\Gamma}}_p^{M_1} + \frac{n_1}{n_1+n_2}\widehat{\boldsymbol{\Gamma}}_p^{M_2}\right)^{-1}\left(\widehat{\boldsymbol{p}}^{(1)} - \widehat{\boldsymbol{p}}^{(2)}\right) \tag{23}$$

$$W^\theta_{n_1,n_2} = \frac{n_1 n_2}{n_1+n_2}\left(\widehat{\boldsymbol{\theta}}^{(1)} - \widehat{\boldsymbol{\theta}}^{(2)}\right)^\top \left(\frac{n_2}{n_1+n_2}\widehat{\boldsymbol{\Gamma}}_\theta^{M_1} + \frac{n_1}{n_1+n_2}\widehat{\boldsymbol{\Gamma}}_\theta^{M_2}\right)^{-1}\left(\widehat{\boldsymbol{\theta}}^{(1)} - \widehat{\boldsymbol{\theta}}^{(2)}\right). \tag{24}$$

If the hypotheses of Theorem 4.2 hold and $H_0^p:\ \mathbf{p}_1 = \mathbf{p}_2$ is true, then

$$W^p_{n_1,n_2} \rightsquigarrow \chi^2_{DoF}$$

with $DoF = D(D-2)$ when there is no absorbing state and $DoF = (D-1)(D-2)$ when there is an absorbing state. Similarly, if $H_0^\theta:\ \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$ is true, then

$$W^\theta_{n_1,n_2} \rightsquigarrow \chi^2_{DoF}$$

with $DoF = D(D-1)d$ when there is no absorbing state and $DoF = (D-1)^2 d$ when there is an absorbing state.

4.2.2 Two-sample likelihood ratio tests

We consider another approach based on the following likelihood ratio,

$$\lambda_{n_1,n_2} = \frac{\prod_{\ell=1}^{n_1}\mathcal{L}(\boldsymbol{S}^1_\ell, \widehat{\mathbf{p}}, \widehat{\boldsymbol{\theta}})\prod_{\ell=1}^{n_2}\mathcal{L}(\boldsymbol{S}^2_\ell, \widehat{\mathbf{p}}, \widehat{\boldsymbol{\theta}})}{\prod_{\ell=1}^{n_1}\mathcal{L}(\boldsymbol{S}^1_\ell, \widehat{\mathbf{p}}^1, \widehat{\boldsymbol{\theta}}^1)\prod_{\ell=1}^{n_2}\mathcal{L}(\boldsymbol{S}^2_\ell, \widehat{\mathbf{p}}^2, \widehat{\boldsymbol{\theta}}^2)} \tag{25}$$

where $(\widehat{\mathbf{p}}^1, \widehat{\boldsymbol{\theta}}^1)$ and $(\widehat{\mathbf{p}}^2, \widehat{\boldsymbol{\theta}}^2)$ are the maximum likelihood estimators of $(\mathbf{p}, \boldsymbol{\theta})$ based on the first and second sample of trajectories and $(\widehat{\mathbf{p}}, \widehat{\boldsymbol{\theta}})$ are the maximum likelihood estimators under H_0. The asymptotic distribution of the likelihood ratio is given in the theorem below when the two samples of trajectories are drawn according to the same SMP.

Theorem 4.3. *Suppose that the conditions of Theorem 3.2 are in force and* $\frac{n_2}{n_1+n_2} \to f \in (0,1)$*, as* n_1 *and* n_2 *tend to infinity. If the null hypothesis* $H_0:\ (\mathbf{p}^1, \boldsymbol{\theta}^1) = (\mathbf{p}^2, \boldsymbol{\theta}^2)$ *is true then*

$$\lambda_{n_1,n_2} \rightsquigarrow \chi^2_{DoF}$$

where the degrees of freedom are equal to $DoF = D(D-2) + D(D-1)d$ *when there is no absorbing state and* $DoF = (D-1)(D-2) + (D-1)^2 d$ *when there is an absorbing state.*

The proof of Theorem 4.3, given in the Appendix, has many similarities to the proof of Theorem 1 in Dannemann and Holzmann (2008), which studies likelihood ratio tests for two samples of hidden Markov models. As in Dannemann and Holzmann (2008), it is possible to consider a more general null hypothesis relying on a regular r dimensional restriction $H_0 : R(\mathbf{p}^1, \boldsymbol{\theta}^1, \mathbf{p}^2, \boldsymbol{\theta}^2) = 0$, provided that the constrained maximum likelihood estimator is consistent.

5 A simulation study for two-sample tests

A small simulation study is conducted to evaluate and compare the effectiveness of the two-sample Wald type test and likelihood ratio test under different scenarios, with varying sample sizes and number of states.

5.1 Experimental protocol

To generate trajectories of an SMP, we must have its initial probability, its transition matrix and the distribution of the sojourn times.

As in Lecuelle et al. (2018) and Cardot et al. (2019), we consider that the sojourn time distribution only depends on the current state i, meaning that $\theta_{ij} = \theta_i$, for $j \neq i$. We also suppose that the distribution for the sojourn times is a Gamma distribution. Indeed the Gamma distribution provides realistic approximations to the sojourn time distributions for sensory analysis data (see Frascolla et al. (2022) for a comparison of the fit of different parametric distributions for sojourn times). Consequently, with these simplifications, the vector of estimated parameters $\widehat{\boldsymbol{\theta}}$ is of dimension $2D$ and the block diagonal matrix $\boldsymbol{\Gamma}_{\theta_0}^M$ is of dimension $2D \times 2D$ with diagonal terms $\boldsymbol{\Gamma}_{\theta_i}^M$, that can be expressed as follows,

$$\begin{aligned}\boldsymbol{\Gamma}_{\theta_i}^M &= \frac{1}{\mathbb{E}\left[N_i^M\right]} \left(-\mathbb{E}\left[\nabla_{\theta_i\theta_i} \ln\left(f(X, \theta_{0_i})\right)\right]\right)^{-1} \\ &= \frac{1}{\mathbb{E}\left[N_i^M\right]} \begin{pmatrix} trigamma(a_i) & -\frac{1}{\lambda_i} \\ -\frac{1}{\lambda_i} & \frac{a_i}{\lambda_i^2} \end{pmatrix}^{-1}\end{aligned}$$

where $a_i > 0$ and $\lambda_i > 0$ are the parameters characterizing the Gamma distribution, with density function $f(x, a, \lambda) = \frac{x^{a-1}\lambda^a \exp(-\lambda x)}{\Gamma(a)}$, for $x \geq 0$, where $\Gamma(a)$ is the Gamma function. In our experiments, as in Frascolla et al. (2022), the values of a_i and λ_i are given by the estimated values on a real sensory analysis data set and the initial probabilities (resp. the rows of the transition matrices) are obtained by normalizing by their sum $D-1$ (resp. $D-2$) independent uniform random variables on $[0, 1]$.

To evaluate the two-sample tests performances according to how distant are the distributions of the two semi-Markov processes, we first define two transition matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, as well as two sets of parameters for the sojourn time distributions, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. We then define, for $\varepsilon \in [0, 1]$, $\mathbf{P}_\varepsilon = (1-\varepsilon)\mathbf{P}_1 + \varepsilon\mathbf{P}_2$ and $\boldsymbol{\theta}_\varepsilon = (1-\varepsilon)\boldsymbol{\theta}_1 + \varepsilon\boldsymbol{\theta}_2$.

We generate n_1 trajectories of an SMP having parameters $(\mathbf{P}_1, \boldsymbol{\theta}_1)$ and n_2 trajectories of an SMP with parameters $(\mathbf{P}_\varepsilon, \boldsymbol{\theta}_\varepsilon)$, for different values of ε. When $\varepsilon = 0$, the two distributions are the same and the rejection level of the two-sample tests of equality should be equal to the nominal level. As ε increases, the two different distributions become more and more distant and the power of the tests is expected to increase.

All the trajectories are observed over $M = 5$ transitions and we consider different

sample sizes, ranging from $n_1 = n_2 = 100$ to $n_1 = n_2 = 800$ as well as different number of states, $D = 4, 7$ or $D = 10$.

5.2 Simulation results

The nominal level of the tests is chosen to be equal to 0.05, so that the null hypothesis of equality in distribution is rejected for the likelihood ratio test (respectively for the Wald type test) when λ_{n_1,n_2} (resp. $W^{p,\theta}_{n_1,n_2}$) is larger than the quantile of order 0.95 of a χ^2 distribution with $D(D-2)+2D$ degrees of freedom. To get a good approximation to the true level and power of the tests, each procedure is repeated 1000 times.

Number of states	$D = 4$	$D = 10$
With $n_1 = n_2 = 100$ trajectories		
LR test	5.1	15.3
Wald	5.5	8.9
With $n_1 = n_2 = 200$ trajectories		
LR test	5.8	10.2
Wald	5.2	5.6
With $n_1 = n_2 = 500$ trajectories		
LR test	4.5	5.4
Wald	4.7	4.8

Table 1: Empirical levels (when $\varepsilon = 0$), with a nominal rejection rate of 5%, for the two-sample tests with $D = 4$ and $D = 10$ states and sample sizes $n_1 = n_2 = 100$, $n_1 = n_2 = 200$, $n_1 = n_2 = 500$.

We first study the distribution of the test statistics under the null hypothesis and consider $n_1 = n_2 = 200$ trajectories and $D = 4$ states, $n_1 = n_2 = 600$ trajectories and $D = 7$ states, $n_1 = n_2 = 800$ trajectories and $D = 10$ states with $M = 5$ transitions. The histograms obtained for the Wald type test are drawn in Figure 1 and those for the likelihood ratio test in Figure 2. The χ^2 distribution is represented in red in the two Figures. The distribution of the p-values are also represented in these two Figures and compared with the uniform distribution on $[0, 1]$. These Figures confirm that the distribution of the two test statistics is well approximated by a χ^2 distribution in the considered cases.

We present in Table 1 the empirical rejection rate, for an expected nominal level of 5% in the different scenarios. For small samples ($n_1 = n_2 = 100$), the two test have an empirical rejection rate close to the expected one only when the number of states is small ($D = 4$). When $D = 10$, the LR test over reject the null hypothesis even if the sample size gets larger ($n_1 = n_2 = 200$). As expected the two tests have similar empirical level when n_1 and n_2 are large. Overall, the Wald approach seems to be more reliable under the null

hypothesis.

Approximated rejection rates when the null hypothesis is not true, considering alternative hypothesis controlled by ε, in are given in Figure 3. The Wald approach seems to be slightly more powerful when the sample sizes are not large ($n_1 = n_2 = 100$) and the number of states is small $D = 4$. When $D = 10$, the LR test which rejects two often the null hypothesis under H_0 cannot be reliable with sample sizes $n_1 = n_2 = 200$. the number of or when the number D of states is not small ($D = 10$). Finally, when the sample size is large ($n_1 = n_2 = 500$) the two testing procedures have similar performances.

6 Concluding remarks

We have studied in this work the asymptotic distribution of two-sample tests of equality in law of semi-Markov processes with parametric sojourn time distributions based on the likelihood ratio statistics and a Wald type procedure. It has been confirmed with a small simulation study that when the sample sizes are sufficiently large both approaches are effective, with similar performances. However, when the experiments do not correspond to the asymptotic regime, that is to say when the sample sizes are too small or the number D of states are large, it seems that the Wald type approach behaves better.

There are many directions that deserve further investigation. When D is large, it would certainly be interesting to study the behavior of two-sample test statistics adopting a sparse point of view. Several works have been made recently in that sparse context to test equality of the law of categorical variables (see Dette and Dörnemann (2020) for a study the behavior of likelihood ratio tests and Plunkett and Park (2019) for new test statistics). Another interesting direction would be to consider, in a two-sample framework, a bayesian point of view (see Votsi et al. (2021) for a recent work in a general framework of semi-Markov processes).

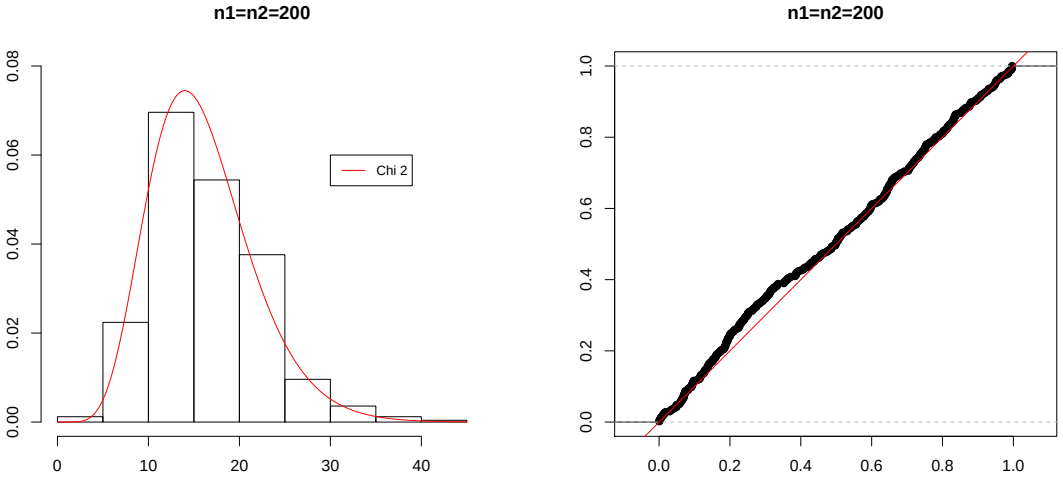
Acknowledgements Calculations were performed using HPC resources from DNUM CCUB (Centre de Calcul de l'Université de Bourgogne). Cindy Frascolla's doctoral thesis is financially supported by the Bourgogne—Franche Comté Regional Council.

References

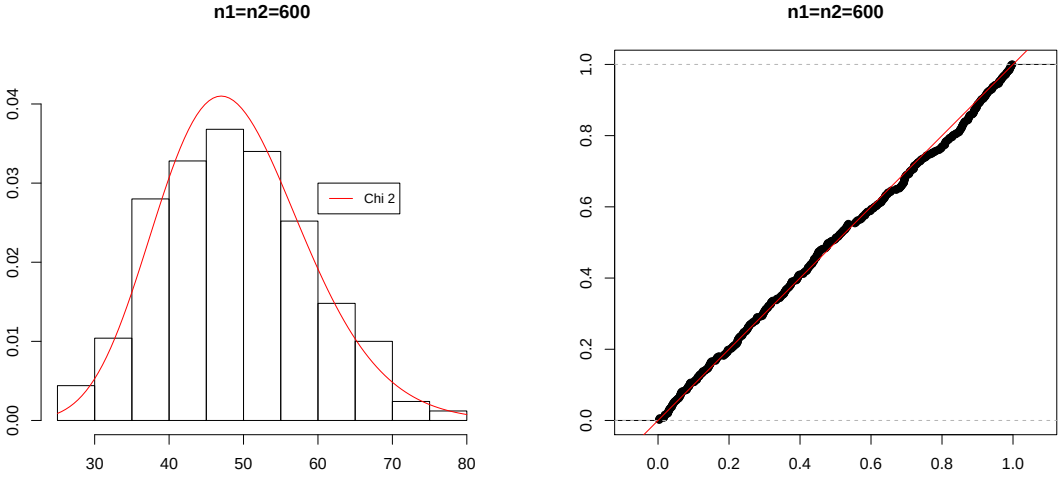
Agresti, A. (2002). *Categorical data analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], New York, second edition.

Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *Ann. Math. Statist.*, 28:89–110.

$D = 4$ states and 16 *ddl*



$D = 7$ states and 49 *ddl*



$D = 10$ states and 100 *ddl*

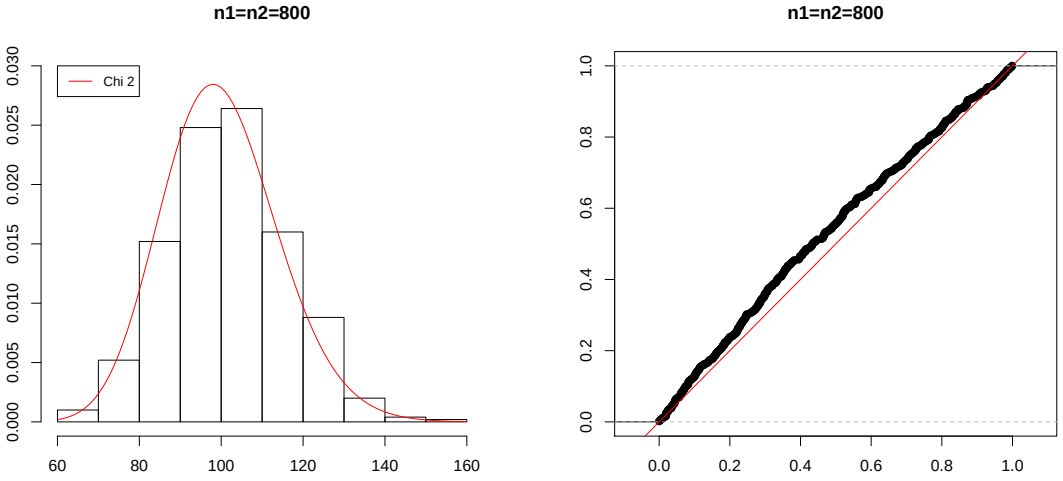
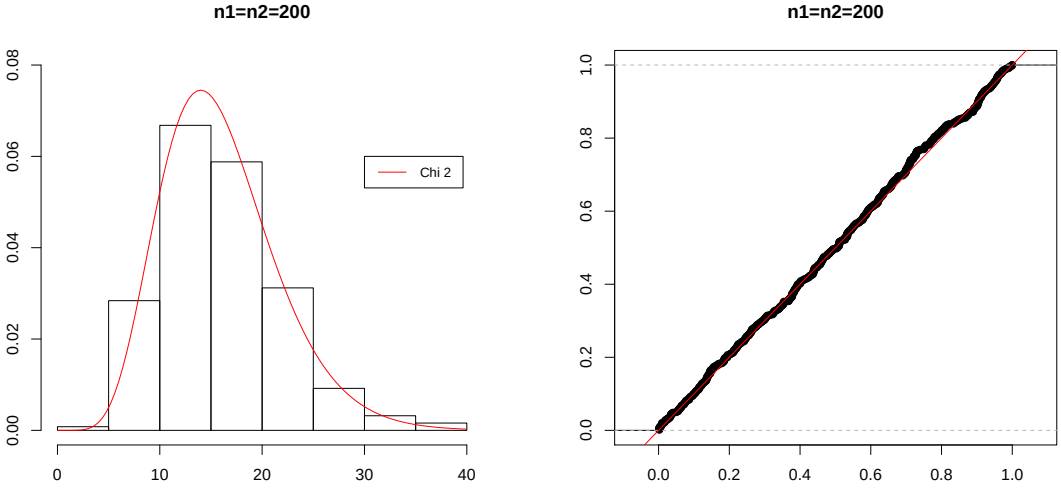
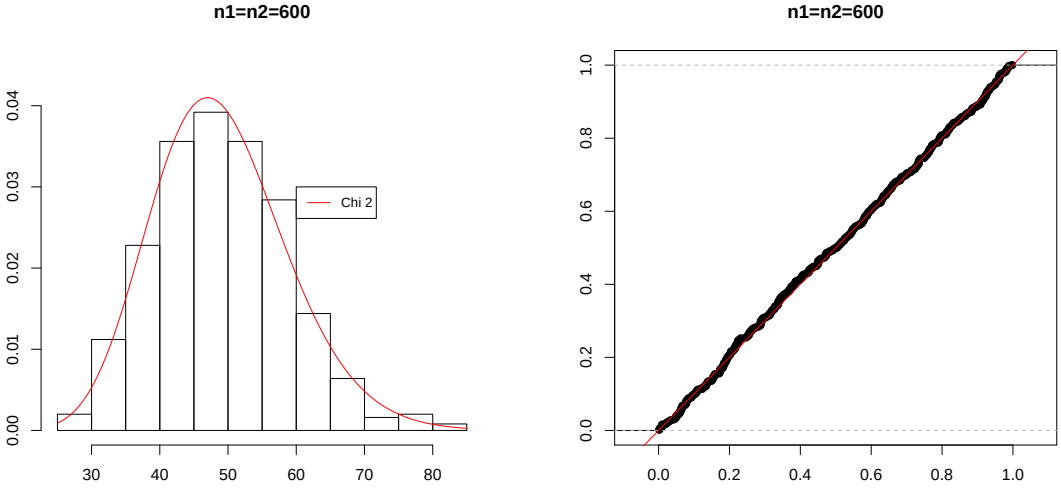


Figure 1: Histogram of the test statistics (left) and Ecdf of the p-values compared with the uniform distribution on $[0, 1]$ (right) for $D = 4$, $D = 7$ and $D = 10$ states (Wald type test).

$D = 4$ states and 16 *ddl*



$D = 7$ states and 49 *ddl*



$D = 10$ states and 100 *ddl*

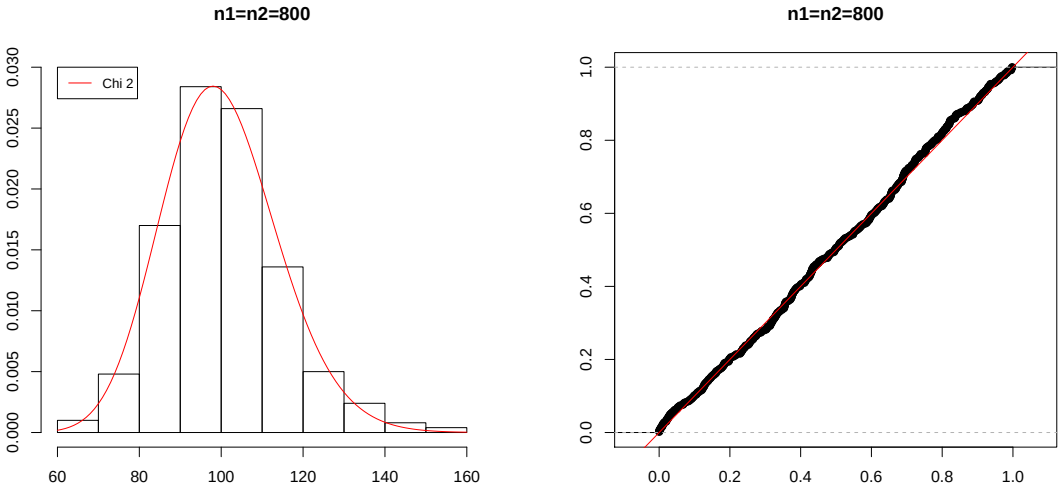


Figure 2: Histogram of the test statistics (left) and Ecdf of the p-values compared with the uniform distribution on $[0, 1]$ (right) for $D = 4$, $D = 7$ and $D = 10$ states (Likelihood ratio test).

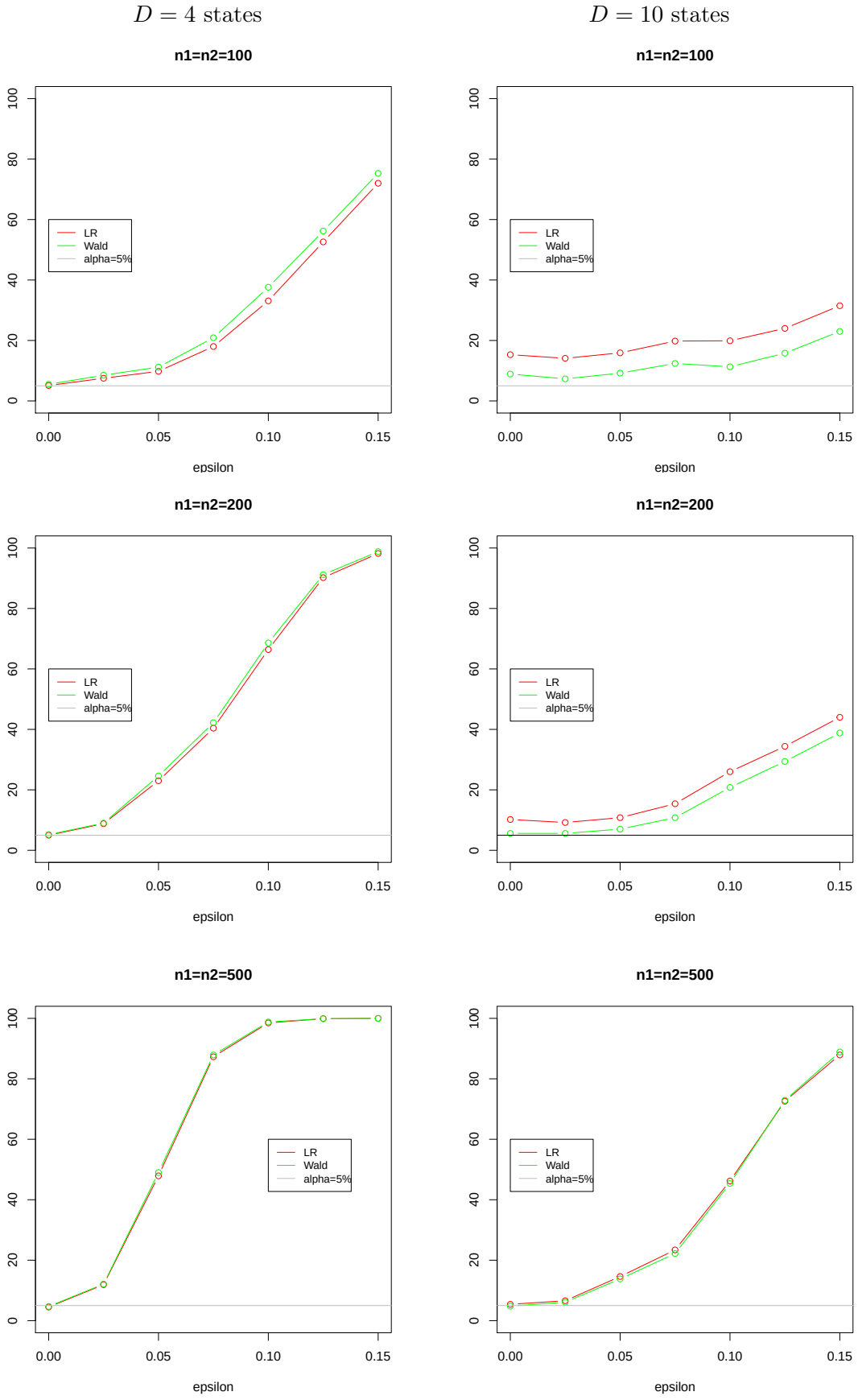


Figure 3: Empirical level (when $\epsilon = 0$) and power of the two-sample tests for $D = 4$ and $D = 10$ states and sample sizes $n_1 = n_2 = 100$, $n_1 = n_2 = 200$, $n_1 = n_2 = 500$.

Barbu, V. S., Karagrigoriou, A., and Makrides, A. (2017). Semi-Markov modelling for multi-state systems. *Methodol. Comput. Appl. Probab.*, 19(4):1011–1028.

Barbu, V. S. and Limnios, N. (2008). *Semi-Markov chains and hidden semi-Markov models toward applications : their use in reliability and DNA analysis.* Springer Science + Business Media, New York.

Billingsley, P. (1961). *Statistical inference for Markov processes.* Statistical Research Monographs, Vol. II. The University of Chicago Press, Chicago, Ill.

Cardot, H., Lecuelle, G., Schlich, P., and Visalli, M. (2019). Estimating finite mixtures of semi-Markov chains: an application to the segmentation of temporal sensory data. *J. R. Stat. Soc. Ser. C. Appl. Stat.*, 68(5):1281–1303.

Dannemann, J. and Holzmann, H. (2008). The likelihood ratio test for hidden markov models in two-sample problems. *Computational Statistics and Data Analysis*, 52(4):1850–1859.

Dette, H. and Dörnemann, N. (2020). Likelihood ratio tests for many groups in high dimensions. *J. Multivariate Anal.*, 178:104605, 16.

Ferguson, T. S. (1996). *A course in large sample theory.* Texts in Statistical Science Series. Chapman & Hall, London.

Frascolla, C., Lecuelle, G., Schlich, P., and Cardot, H. (2022). Two sample tests for semi-Markov processes with parametric sojourn time distributions: an application in sensory analysis. *Computational Statistics*, 37:2553–2580.

Gut, A. (2009). *Stopped random walks.* Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition.

Hjort, N. L. and Pollard, D. (2011). Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806.*

Kemeny, J. G. and Snell, J. L. (1976). *Finite Markov chains.* Springer-Verlag, New York-Heidelberg.

Lecuelle, G., Visalli, M., Cardot, H., and Schlich, P. (2018). Modeling temporal dominance of sensations with semi-Markov chains. *Food Quality and Preference*, 67:59–66.

Lehmann, E. L. and Romano, J. P. (2005). *Testing statistical hypotheses.* Springer Texts in Statistics. Springer, New York, third edition.

Limnios, N. and Oprişan, G. (2001). *Semi-Markov processes and reliability.* Statistics for Industry and Technology. Birkhäuser Boston, Inc., Boston, MA.

Moore, E. H. and Pyke, R. (1968). Estimation of the transition distributions of a Markov renewal process. *Ann. Inst. Statist. Math.*, 20:411–424.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of econometrics, Vol. IV*, volume 2 of *Handbooks in Econom.*, pages 2111–2245. North-Holland, Amsterdam.

Plunkett, A. and Park, J. (2019). Two-sample test for sparse high-dimensional multinomial distributions. *TEST*, 28(3):804–826.

Pons, O. (2006). Semi-parametric estimation for a semi-Markov process with left-truncated and right-censored observations. *Statistics & Probability Letters*, 76:952–958.

Trevezas, S. and Limnios, N. (2011). Exact MLE and asymptotic properties for nonparametric semi-Markov models. *J. Nonparametr. Stat.*, 23(3):719–739.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Votsi, I., Gayraud, G., Barbu, V., and Limnios (2021). Hypotheses testing and posterior concentration rates for semi-Markov processes. *Statistical Inference for Stochastic Processes*, 24:707–732.

Appendix

Before giving the proofs of the theorems, we recall some useful results.

Some useful results

We recall Wald's identity which is often used in this work (see for example Gut (2009), Theorem 5.3, Chapter 1 for a proof). Consider *i.i.d* random vectors $(V_k, W_k), k \geq 1$ and a stopping time N, with respect to an increasing sequence of sub-σ-algebras $\mathcal{F}_k, k \geq 1$ such that (V_k, W_k) is $\mathcal{F}_k$ measurable and independent of $\mathcal{F}_{k-1}$.

Theorem 6.1. *(Wald's identity)*
Suppose that $\mathbb{E}(V_k) = \mu_v$ *and* $\mathbb{E}(N) < \infty$. *Then*

$$\mathbb{E}\left[\sum_{k=1}^{N} V_k\right] = \mu_v \mathbb{E}(N).$$

If (V_k, W_k) *have finite variance, and* $\mathbb{E}(W_k) = \mu_w$ *then*

$$\mathbb{E}\left[\left(\sum_{k=1}^{N} V_k - N\mu_v\right)\left(\sum_{k=1}^{N} W_k - N\mu_w\right)\right] = \mathbb{C}ov(V_1, W_1)\mathbb{E}(N).$$

We also recall some basic properties of the moments of the counting processes.

Lemma 6.1. *If* $\mathbf{A}_1$ *and* $\mathbf{A}_2$ *hold and* $\mathbb{E}[M] < +\infty$ *then* $0 < \mathbb{E}\left[N_i^M\right] < +\infty$ *and*

$$\mathbb{E}\left[N_i^M\right] = \sum_{m=1}^{+\infty}\left(\sum_{k=0}^{m-1}\sum_{j\in E}\alpha_{0_j} p_{0_{ji}}^{(k)}\right)\mathbb{P}\left(M = m\right)$$
$$\mathbb{E}\left[N_{ij}^M\right] = p_{0_{ij}}\mathbb{E}\left[N_i^M\right].$$

The following Lemma is a direct adaptation of Theorem 3.3.5 in Kemeny and Snell (1976)).

Lemma 6.2. *For* $i \neq j \in \{1, \ldots, D-1\}$,

$$\mathbb{E}\left[M_D | J_0 = i\right] = \sum_{j=1}^{D-1} F_{ij}$$
$$Var[M_D | J_0 = i] = 2\sum_{j=1}^{D-1} F_{ij}\mathbb{E}\left[M_D | J_0 = j\right] - \mathbb{E}\left[M_D | J_0 = i\right] - \left(\mathbb{E}\left[\tau | J_0 = i\right]\right)^2$$
$$\mathbb{E}\left[N_j^{M_D}\right] = \sum_{i=1}^{D-1} F_{ij}\alpha_{0_i}$$
$$\mathbb{E}\left[N_{ij}^{M_D}\right] = p_{0_{ij}}\mathbb{E}\left[N_j^{M_D}\right].$$

Thus $\mathbb{E}\left[M_D\right] = \sum_{i=1}^{D-1}\mathbb{E}\left[M_D | J_0 = i\right]\alpha_{0_i}$, as well as finite variance (see Corollary 3.3.6 in Kemeny and Snell (1976)).

Proofs

Proof. of Theorem 3.1.

The transition probabilities. *The almost sure consistency of* $\widehat{\mathbf{P}}$ *is a direct consequence of the strong law of large numbers and Wald's identity.*

Define $N_i^M = \sum_{k=0}^{M-1} \mathbf{1}_{\{J_k=i\}}$ *the number of visits of state* i *during the first* $M-1$ *transitions and note that we can rewrite* $N_{ij}^M = \sum_{k=0}^{M-1} \mathbf{1}_{\{J_k=i;J_{k+1}=j\}}$*, as follows*

$$N_{ij}^M = \sum_{k=1}^{N_i^M} B_{ij}^k,$$

where B_{ij}^k *is a Bernoulli variable taking value one if after the* k*th visit of state* i*, the next visited state is state* j *and zero else. By the Markov property of* $(J_k)_{k\geq 0}$*, the variables* $(B_{ij}^k)_{k\geq 1}$ *are independent with expected value* $\mathbb{P}(B_{ij}^k = 1) = p_{0_{ij}}$*. Furthermore* $0 < \mathbb{E}\left(N_i^M\right) \leq \mathbb{E}(M) < +\infty$*, so that we deduce with Wald's identity that*

$$\mathbb{E}[N_{ij}^M] = p_{0_{ij}} \mathbb{E}\left[N_i^M\right], \tag{26}$$

and $\mathbb{E}[N_{ij}^M] > 0$ *with assumption* $\mathbf{A}_2$*. Since the trajectories are supposed to be independent, the strong law of large numbers gives us that* $n^{-1}\sum_{\ell=1}^n N_{ij}^{(\ell)} \to \mathbb{E}[N_{ij}^M] = p_{0,ij}\mathbb{E}\left[N_i^M\right]$ *and* $n^{-1}\sum_{\ell=1}^n N_i^{(\ell)} \to \mathbb{E}\left[N_i^M\right] > 0$ *almost surely, as* $n \to \infty$*. We get with the continuous mapping theorem (see van der Vaart (1998), Theorem 2.3) that* $\widehat{p}_{ij}$ *converges almost surely to* $\mathbb{E}[N_{ij}^M]/\mathbb{E}\left[N_i^M\right] = p_{0_{ij}}$ *and the announced result. Note that when* $M = M_D$*, we have with Theorem 3.3.5 in Kemeny and Snell (1976),* $\mathbb{E}\left[N_i^{M_D}\right] = \sum_{j=1}^{D-1} F_{ji}\alpha_{0_j}$*.*

The sojourn time parameters. *By the renewal property of the semi-Markov process, the sojourn times in state* $\{i\}$ *when the next state is* $\{j\}$ *form a sequence* $(X_{ij}^k)_{k\geq 1}$ *of i.i.d random variables, with density* $f(.,\theta_{0_{ij}})$*. We get with Wald's identity,*

$$\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\sum_{k=1}^{N_{ij}^M} \ln\left(f\left(X_{ij}^k,\theta_{ij}\right)\right)\right] = \mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_{ij}^M\right]\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\ln(f(X_{ij},\theta_{ij}))\right]$$

where X_{ij} *is a random variable with density* $f(.,\theta_{0_{ij}})$*. Define* $Q_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})$ *to be the expectation of* $\widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$*, with respect to the distribution characterized by parameter* $\boldsymbol{\theta}_0$*, we have*

$$\begin{aligned} Q_{\boldsymbol{\theta}_0}(\boldsymbol{\theta}) &= \mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right] \\ &= \sum_{\substack{i,j\in E \\ j\neq i}} \mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_{ij}^M\right]\mathbb{E}_{\boldsymbol{\theta}_0}\left[\ln(f(X_{ij},\theta_{ij}))\right]. \end{aligned} \tag{27}$$

We deduce from $\mathbf{A}_4$ *and* $\mathbf{A}_5$ *and the information inequality given in Lemma 2.2 in Newey and McFadden (1994) that* $\mathbb{E}_{\boldsymbol{\theta}_0}\left[\ln(f(X_{ij},\theta_{ij}))\right]$ *has a unique maximum at* $\theta_{ij} = \theta_{0_{ij}}$*. Since*

$\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_{ij}^M\right] > 0$ *and by the additivity of* $Q_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})$ *given in (27) we get that* $Q_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})$ *attains its unique maximum at* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. *We can note that* $Q_{\boldsymbol{\theta}_0}$ *is continuous since it is the sum of continuous functions.*

By assumption $\mathbf{A}_5$, $\sum_{\substack{i,j\in E\\ j\neq i}}\sum_{k=1}^{N_{ij}^M}\ln\left(f\left(X_{ij}^{k,l},\theta_{ij}\right)\right)$ *is continuous at each* $\theta \in \boldsymbol{\Theta}$ *with probability one, and we have*

$$\begin{aligned}\left|\sum_{\substack{i,j\in E\\ j\neq i}}\sum_{k=1}^{N_{ij}^M}\ln\left(f\left(X_{ij}^{k,l},\theta_{ij}\right)\right)\right| &\leq \sum_{\substack{i,j\in E\\ j\neq i}}\sum_{k=1}^{N_{ij}^M}\left|\ln\left(f\left(X_{ij}^{k,l},\theta_{ij}\right)\right)\right| \\ &\leq \sum_{\substack{i,j\in E\\ j\neq i}}\sum_{k=1}^{N_{ij}^M} g(X_{ij}^{k,l})\end{aligned}$$

with $g(X)$ *defined as* $\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\ln f(X,\theta)|$ *which is reached as* $\boldsymbol{\Theta}$ *is compact and the function* $\ln f(X,\theta)$ *is continuous in* θ *(assumption* $\mathbf{A}_5$*). We get with Wald's identity that*

$$\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\sum_{\substack{i,j\in E\\ j\neq i}}\sum_{k=1}^{N_{ij}^M} g(X_{ij}^{k,l})\right] = \sum_{\substack{i,j\in E\\ j\neq i}}\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_{ij}^M\right]\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[g(X_{ij})\right].$$

Thanks to assumption $\mathbf{A}_5$, $\mathbb{E}_{\theta_0}\left[g(X)\right] < \infty$ *and by hypothesis* $\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_{ij}^M\right] < \infty$. *Thus, we have* $\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\sum_{\substack{i,j\in E\\ j\neq i}}\sum_{k=1}^{N_{ij}^M} g(X_{ij}^{k,l})\right] < \infty$.

The independence of the n trajectories, the additional compactness assumption $\mathbf{A}_3$ *and Lemma 2.4 in Newey and McFadden (1994) allows to obtain a uniform law of large numbers,*

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) - Q_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})\right| \to 0 \quad \textit{almost surely when } n \textit{ tends to infinity.}$$

We can finally deduce from Theorem 2.1 in Newey and McFadden (1994) that $\widehat{\boldsymbol{\theta}}$ *converges almost surely to* $\boldsymbol{\theta}_0$.

□

Proof. of Theorem 3.2.

The case with an absorbing state $\{D\}$ *follows the same lines as the proof without any absorbing state and is not given. First note that by the additive property of the average likelihood criterion (7) the covariance between* $\sqrt{n}(\widehat{\mathbf{p}} - \mathbf{p_0})$ *and* $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ *is equal to zero. We first derive the asymptotic normality of* $\widehat{\mathbf{p}}$.

The transition probabilities. *The proof is similar to the proof of Lemma 4.2 in Trevezas and Limnios (2011) and relies on the multivariate central limit theorem applied on the elements of the empirical score,* $\nabla_{\mathbf{P}}\widehat{Q}_{\mathbf{P}}(\mathbf{P})$ *which are i.i.d, with expectation zero at* $\mathbf{p} = \mathbf{p}_0$.

As n tends to infinity, we have

$$\sqrt{n}\nabla_{\boldsymbol{p}}\widehat{Q}_{\boldsymbol{p}}(\boldsymbol{p}_0) \rightsquigarrow \mathcal{N}(0, \boldsymbol{\Delta}_p)$$

where the covariance terms are defined as follows

$$\boldsymbol{\Delta}_{p_{ij,ab}} = \mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\left(\frac{N_{ij}^M}{p_{0_{ij}}} - N_i^M\right)\left(\frac{N_{ab}^M}{p_{0_{ab}}} - N_a^M\right)\right].$$

Conditioning on N_i^M we get that $\boldsymbol{\Delta}_{p_{ij,ab}} = 0$ if $a \neq i$ so that $\boldsymbol{\Delta}_p$ is block diagonal. Furthermore, given N_i^M, the vector with components N_{ij}^M for $j \neq i$ has a multinomial distribution, with probability of success $p_{0_{ij}}$. We thus obtain that $\boldsymbol{\Delta}_{p_{ij,ij}} = \left(\frac{1}{p_{0ij}} - 1\right)\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_i^M\right]$ and for $b \neq j$, $\boldsymbol{\Delta}_{p_{ij,ib}} = -\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_i^M\right]$.
We have,

$$\nabla_{p_{ab}p_{ij}}\widehat{Q}_{\boldsymbol{p}}(\boldsymbol{p}) = 0$$
$$\nabla_{p_{ij}p_{ij}}\widehat{Q}_{\boldsymbol{p}}(\boldsymbol{p}) = -\frac{1}{n}\sum_{l=1}^{n}\frac{N_{ij}^{\ell}}{p_{ij}^2}.$$

Let $\mathbf{H}^M(\mathbf{p})$ be $\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{p}_0,\boldsymbol{\theta}_0}\left[\nabla_{\mathbf{pp}}\widehat{Q}_{\mathbf{p}}(\mathbf{p})\right]$. Thanks to the weak law of large numbers and the Continuous Mapping Theorem, we obtain that, as $n \to +\infty$, $\nabla_{\boldsymbol{pp}}\widehat{Q}_{\boldsymbol{p}}(\widehat{\boldsymbol{p}}) \to \mathbf{H}^M(\boldsymbol{p}_0)$ in probability so that

$$\sup_{\boldsymbol{p}\in\mathcal{N}}\left\|\nabla_{\boldsymbol{pp}}\widehat{Q}_{\boldsymbol{p}}(\boldsymbol{p}) - \mathbf{H}^M(\boldsymbol{p})\right\| \to 0 \quad \textit{in probability.}$$

Let $\mathbf{H}_p^M$ be $\mathbf{H}^M(\mathbf{p_0})$ which is a diagonal matrix composed of the terms $-\frac{\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_i^M\right]}{p_{0_{ij}}}$. As each term $p_{0_{ij}} > 0$ (condition $\mathbf{A}_2$), the matrix $\mathbf{H}_p^M$ is non singular and its inverse is diagonal and composed of the terms $-\frac{p_{0_{ij}}}{\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_i^M\right]}$.

We get by Theorem 3.1 in Newey and McFadden (1994) that

$$\sqrt{n}\left(\widehat{\mathbf{p}} - \mathbf{p}_0\right) \rightsquigarrow \mathcal{N}\left(0, \left(\mathbf{H}_p^M\right)^{-1}\boldsymbol{\Delta}_{\boldsymbol{p}}\left(\mathbf{H}_p^M\right)^{-1}\right).$$

Simple calculations enable to obtain that $\left(\mathbf{H}_p^M\right)^{-1}\boldsymbol{\Delta}_{\boldsymbol{p}}\left(\mathbf{H}_p^M\right)^{-1}$ is equal to $\boldsymbol{\Gamma}_{p_0}^M$.

The sojourn time parameters. *Consider the partial derivatives of $\widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ and denote by $\nabla_{\theta_{ij}}\widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ the gradient vector partial derivatives with respect to the components of θ_{ij}:*

$$\nabla_{\theta_{ij}}\widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \left(\frac{\partial\widehat{Q}_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}_{ij,1}}, \ldots, \frac{\partial\widehat{Q}_{\boldsymbol{\theta}}}{\partial\boldsymbol{\theta}_{ij,d}}\right) \in \mathbb{R}^d$$

for $i = 1, \ldots, D$ and $j = 1, \ldots, D$ and $j \neq i$. We have

$$\nabla_{\theta_{ij}} \widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \frac{1}{n} \sum_{\ell=1}^{n} \sum_{k=1}^{N_{ij}^{(\ell)}} \nabla_{\theta_{ij}} \ln \left(f(x_{ij}^{(\ell,k)}, \theta_{ij}) \right). \tag{28}$$

Taking the expectation, we get with Wald's identity that

$$\mathbb{E}\left[\nabla_{\boldsymbol{\theta}_{ij}} \widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right] = \mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_{ij}^M\right] \mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\nabla_{\theta_{ij}} \ln\left(f(X_{ij};\theta_{ij})\right)\right]$$

so that, when $\theta_{ij} = \theta_{0_{ij}}$ and hypothesis $\mathbf{A}_6$ ((iv to vi) which allows to interchange the order of integration and differentiation) is true, $\mathbb{E}\left[\nabla_{\boldsymbol{\theta}_{ij}} \widehat{Q}_{\boldsymbol{\theta}_0}(\boldsymbol{\theta})\right] = 0$. The variance and covariance terms can be calculated by applying Wald's identity for the covariance terms (see Theorem 6.1 in the Appendix). We get

$$\begin{aligned}
\boldsymbol{\Delta}_{\theta_{ij}} &= \mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\left(\sum_{k=1}^{N_{ij}^M} \nabla_{\theta_{ij}} \ln\left(f(X_{ij}^k, \theta_{0_{ij}})\right)\right)\left(\sum_{k=1}^{N_{ij}^M} \nabla_{\theta_{ij}} \ln\left(f(X_{ij}^k, \theta_{0_{ij}})\right)\right)^\top\right] \\
&= \mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_{ij}^M\right] \mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\nabla_{\theta_{ij}} \ln\left(f(X_{ij}, \theta_{0_{ij}})\right) \nabla_{\theta_{ij}} \ln\left(f(X_{ij}, \theta_{0_{ij}})\right)^\top\right],
\end{aligned}$$

which is a full rank $d \times d$ matrix under assumption $\mathbf{A}_6$ (v) and since $\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_{ij}^M\right] = p_{0_{ij}} \mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[N_i^M\right] > 0$. It can be shown that the covariance terms, for $(i,j) \neq (a,b)$ are equal to zero by conditioning on N_{ij}^M and N_{ab}^M and the fact that $\ln\left(f(X_{ij}, \theta_{0_{ij}})\right)$ and $\ln\left(f(X_{ab}, \theta_{0_{ab}})\right)$ are independent centered random variables,

$$\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\nabla_{\theta_{ij}} \ln\left(f(X_{ij}, \theta_{0_{ij}})\right) \nabla_{\theta_{ab}} \ln\left(f(X_{ab}, \theta_{0_{ab}})\right)^\top\right] = 0.$$

We thus get with the multivariate central limit theorem that

$$\sqrt{n} \nabla_{\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0) \rightsquigarrow \mathcal{N}\left(0, \boldsymbol{\Delta}_{\boldsymbol{\theta}_0}\right) \tag{29}$$

where $\boldsymbol{\Delta}_{\boldsymbol{\theta}_0}$ is a block diagonal matrix, with diagonal elements $\boldsymbol{\Delta}_{\theta_{ij}}, i = 1, \ldots, D, j \neq i$. Each sub-matrix $\boldsymbol{\Delta}_{\theta_{ij}}$ being non singular, matrix $\boldsymbol{\Delta}$ admits an inverse, which is also a block diagonal matrix, with diagonal elements $\left(\boldsymbol{\Delta}_{\theta_{ij}}\right)^{-1}$.

Note now that condition $\mathbf{A}_6$ (iv to vi) which allows to interchange the order of integration and differentiation implies that (see Theorem 3.3 in Newey and McFadden (1994)),

$$\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\nabla_{\theta_{ij}} \ln\left(f(X_{ij}, \theta_{0_{ij}})\right) \nabla_{\theta_{ij}} \ln\left(f(X_{ij}, \theta_{0_{ij}})\right)^\top\right] = -\mathbb{E}_{\boldsymbol{\alpha}_0,\mathbf{P}_0,\boldsymbol{\theta}_0}\left[\nabla_{\theta_{ij}\theta_{ij}} \ln\left(f(X_{ij}, \theta_{0_{ij}})\right)\right].$$

We denote by $\nabla_{\boldsymbol{\theta}_{ab}\boldsymbol{\theta}_{ij}} \widehat{Q}_{\boldsymbol{\theta}}$ the Hessian matrix with generic elements,

$$\left[\nabla_{\boldsymbol{\theta}_{ab}\boldsymbol{\theta}_{ij}} \widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right]_{u,v} = \frac{\partial^2 \widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_{ij,v} \partial \boldsymbol{\theta}_{ab,u}}, \quad u, v = 1, \ldots, d.$$

It is clear from (28) that if $(i,j) \neq (a,b)$ then $\nabla_{\boldsymbol{\theta}_{ab}\boldsymbol{\theta}_{ij}} \widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = 0$ so that the global Hessian matrix $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is block diagonal.

Following the lines of the Proof of Theorem 3.3 in Newey and McFadden (1994) and using now the set of first order conditions given in (17), we have with the mean value theorem,

$$0 = \nabla_{\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0) + \left[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}})\right] \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \tag{30}$$

where each component of $\overline{\boldsymbol{\theta}}$ belongs to the segment defined by the corresponding components of $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$.

Defining the global neighborhood $\mathcal{N}$ of $\boldsymbol{\theta}$ as the Cartesian product of the neighborhoods $\mathcal{N}_{ij}$ of θ_{ij}, we get as $n \to +\infty$,

$$\sup_{\boldsymbol{\theta} \in \mathcal{N}} \left\| \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) - \mathbf{H}^M(\boldsymbol{\theta}) \right\| \to 0 \quad \textit{in probability}$$

where $\mathbf{H}^M(\boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\alpha}_0, \mathbf{P}_0, \boldsymbol{\theta}_0}\left[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right]$ so that $\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) \to \mathbf{H}^M(\boldsymbol{\theta}_0)$ in probability. Combining finally (29) and (30), we get with Slustky's theorem

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \rightsquigarrow \mathcal{N}\left(0, \left(\mathbf{H}^M(\boldsymbol{\theta}_0)\right)^{-1} \boldsymbol{\Delta}_{\theta_0} \left(\mathbf{H}^M(\boldsymbol{\theta}_0)\right)^{-1}\right)$$

with $\mathbf{H}^M(\boldsymbol{\theta}_0) = -\boldsymbol{\Delta}_{\theta_0}$ and thus $\boldsymbol{\Gamma}^M_{\theta_0} = \left(-\mathbf{H}^M(\boldsymbol{\theta}_0)\right)^{-1}$. This concludes the proof. □

Proof. of Theorem 4.1. *The additive structure of the log likelihood and the fact that the asymptotic variance matrix of the maximum likelihood estimators is block diagonal lead to the following Taylor expansion of $-2\ln\lambda_n$ about $\left(\widehat{\mathbf{p}}, \widehat{\boldsymbol{\theta}}\right)$,*

$$-2\ln\lambda_n = -n\left(\left(\widehat{\boldsymbol{p}} - \boldsymbol{p}_0\right)^\top \nabla_{\boldsymbol{p}\boldsymbol{p}} \widehat{Q}_{\boldsymbol{p}}(\overline{\boldsymbol{p}}) \left(\widehat{\boldsymbol{p}} - \boldsymbol{p}_0\right) + \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^\top \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}) \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)\right). \tag{31}$$

We use also the fact that the maximum likelihood estimators satisfy $\nabla_{\boldsymbol{p}} \widehat{Q}_{\boldsymbol{p}}(\widehat{\mathbf{p}}) = 0$ and $\nabla_{\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) = 0$. The equality holds for some $\overline{\boldsymbol{\theta}}$ (resp. $\overline{\boldsymbol{p}}$) belonging elementwise to the segment defined by $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ (resp. by $\widehat{\boldsymbol{p}}$ and $\boldsymbol{p}_0$). The log-likelihood can be studied independently for the sojourn times and the transition probabilities.

The sojourn time parameters. *Let $-2\ln\lambda_{n,\theta}$ be the part of the test statistics corresponding to the sojourn time parameters:*

$$-2\ln\lambda_{n,\theta} = -n\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^\top \nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}) \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right).$$

Thanks to the weak law of large numbers and the Continuous Mapping Theorem, we have the convergence in probability $-\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}) \to \left(\boldsymbol{\Gamma}^M_{\theta_0}\right)^{-1}$ which means that $-\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \widehat{Q}_{\boldsymbol{\theta}}(\overline{\boldsymbol{\theta}}) = \left(\boldsymbol{\Gamma}^M_{\theta_0}\right)^{-1} + o_p(1)$ (see van der Vaart (1998) Section 2.2 for the definition of notations

$o_p(1)$ *and* $O_p(1)$*). If* H_0 *holds and under the hypotheses of Theorem 3.2 we have that* $\sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right) \rightsquigarrow \mathcal{N}\left(0, \boldsymbol{\Gamma}_{\theta_0}^M\right)$*. Then, we obtain the approximation:*

$$\begin{aligned}-2\ln\lambda_{n,\theta} &= \sqrt{n}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)^\top \left(\boldsymbol{\Gamma}_{\theta_0}^M\right)^{-1}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+o_p(1)\\ &= \left(\sqrt{n}\left(\boldsymbol{\Gamma}_{\theta_0}^M\right)^{-\frac{1}{2}}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\right)^\top \left(\sqrt{n}\left(\boldsymbol{\Gamma}_{\theta_0}^M\right)^{-\frac{1}{2}}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)\right)+o_p(1)\end{aligned}$$

where $\sqrt{n}\left(\boldsymbol{\Gamma}_{\theta_0}^M\right)^{-\frac{1}{2}}\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right) \rightsquigarrow \mathcal{N}\left(0, \boldsymbol{I}\right)$ *with* $\boldsymbol{I}$ *the identity matrix of size* $D(D-1)d$ *if there is no absorbing state and* $(D-1)^2 d$ *else.*

Finally, we obtain:

$$-2\ln\lambda_{n,\theta} \rightsquigarrow \chi^2_{D(D-1)d}$$

when there is no absorbing state and

$$-2\ln\lambda_{n,\theta} \rightsquigarrow \chi^2_{(D-1)^2 d}$$

else.

The transition probabilities. *The idea of the proof is the same as the one for the sojourn times but the difference is due to the fact that the matrix* $\boldsymbol{\Gamma}_{p_0}^M$ *is singular.*

Let $-2\ln\lambda_{n,p}$ *be the part of the test statistics corresponding to the transition probabilities:*

$$-2\ln\lambda_{n,p} = -n\left(\widehat{\boldsymbol{p}}-\boldsymbol{p}_0\right)^\top \nabla_{\boldsymbol{pp}}\widehat{Q}_{\boldsymbol{p}}(\overline{\boldsymbol{p}})\left(\widehat{\boldsymbol{p}}-\boldsymbol{p}_0\right).$$

As for the sojourn times, thanks to the weak law of large numbers and the Continuous Mapping Theorem, we have the convergence in probability $\nabla_{\boldsymbol{pp}}\widehat{Q}_{\boldsymbol{p}}(\overline{\boldsymbol{p}}) \to \boldsymbol{H}_p^M$ *which can be written as* $\nabla_{\boldsymbol{pp}}\widehat{Q}_{\boldsymbol{p}}(\overline{\boldsymbol{p}}) = \boldsymbol{H}_p^M + o_p(1)$*. We have also under the hypotheses of Theorem 3.2 and if* H_0 *holds that* $\sqrt{n}\left(\widehat{\boldsymbol{p}}-\boldsymbol{p}_0\right) \rightsquigarrow \mathcal{N}\left(0, \boldsymbol{\Gamma}_{p_0}^M\right)$*.*
Let $\sqrt{\boldsymbol{p}_{0_i}}^T$ *be the vector of size* $D-1$ *defined as* $\left(\sqrt{p_{0_{i1}}}, \ldots, \sqrt{p_{0_{ij}}}, \ldots, \sqrt{p_{0_{iD}}}\right)$ *with* $j \neq i$ *and let* $\boldsymbol{U}_i$ *be the matrix* $\boldsymbol{I} - \sqrt{\boldsymbol{p}_{0i}}\sqrt{\boldsymbol{p}_{0_i}}^T$ *with* $\boldsymbol{I}$ *the identity matrix of size* $(D-1)$*. Simple calculations enable to show that* $\left(-\boldsymbol{H}_p^M\right)^{\frac{1}{2}}\boldsymbol{\Gamma}_{p_0}^M\left(-\boldsymbol{H}_p^M\right)^{\frac{1}{2}}$ *is a block diagonal matrix of size* $D(D-1)$ *if there is no absorbing state and* $(D-1)^2$ *else with diagonal terms* $\boldsymbol{U}_i$*. It is easy to see that* $\boldsymbol{I} - \sqrt{\boldsymbol{p}_{0i}}\sqrt{\boldsymbol{p}_{0_i}}^T$ *is an orthogonal projection matrix of rank* $D-2$*. Thus, there exists* $\boldsymbol{O}_i$ *an orthogonal matrix such that* $\boldsymbol{\Delta}_i = \boldsymbol{O}_i\left(\boldsymbol{I} - \sqrt{\boldsymbol{p}_{0_i}}\sqrt{\boldsymbol{p}_{0_i}}^T\right)\boldsymbol{O}_i^T$ *with* $\boldsymbol{\Delta}_i$ *the diagonal matrix with* $D-2$ *values of 1 and one 0 for the diagonal. Thus, we have:*

$$\begin{aligned}-2\ln\lambda_{n,p} &= -n\left(\widehat{\boldsymbol{p}}-\boldsymbol{p}_0\right)^\top\left(-\boldsymbol{H}_p^M\right)\left(\widehat{\boldsymbol{p}}-\boldsymbol{p}_0\right)+o_p(1)\\ &= \left(\sqrt{n}\left(-\boldsymbol{H}_p^M\right)^{\frac{1}{2}}\left(\widehat{\boldsymbol{p}}-\boldsymbol{p}_0\right)\right)^\top\left(\sqrt{n}\left(-\boldsymbol{H}_p^M\right)^{\frac{1}{2}}\left(\widehat{\boldsymbol{p}}-\boldsymbol{p}_0\right)\right)+o_p(1)\end{aligned}$$

with $\left(\sqrt{n}\left(-\boldsymbol{H}_p^M\right)^{\frac{1}{2}}\left(\widehat{\boldsymbol{p}}-\boldsymbol{p}_0\right)\right) \rightsquigarrow \mathcal{N}\left(0, \boldsymbol{O}^\top \boldsymbol{\Delta} \boldsymbol{O}\right)$ *where* $\boldsymbol{O}$ *(respectively* $\boldsymbol{\Delta}$*) is the diagonal block matrix composed of the terms* $\boldsymbol{O}_i$ *(respectively* $\boldsymbol{\Delta}_i$*).*

Thus, we obtain that:

$$-2\ln\lambda_{n,p} \rightsquigarrow \chi^2_{rank(\boldsymbol{\Delta})}$$

where $rank(\boldsymbol{\Delta}) = D(D-2)$ *when there is no absorbing state and* $(D-1)(D-2)$ *else.*

The test statistics. *In conclusion, we have, thanks to the asymptotic normality and the block diagonal structure of the asymptotic variance,*

$$\begin{aligned}-2\ln\lambda_n &= -2\ln\lambda_{n,p} - 2\ln\lambda_{n,\theta}\\ &\rightsquigarrow \chi^2_{D(D-2)} + \chi^2_{D(D-1)d} = \chi^2_{D(D-2)+D(D-1)d}\end{aligned}$$

when there is no absorbing state and

$$-2\ln\lambda_n \rightsquigarrow \chi^2_{(D-1)(D-2)} + \chi^2_{(D-1)^2 d} = \chi^2_{(D-1)(D-2)+(D-1)^2 d} \tag{32}$$

when there is an absorbing state and the sequence is observed until absorption.

□

Proof. of Lemma 4.1. *This Lemma is a direct consequence of the asymptotic normality stated in Theorem 3.2 for the maximum likelihood estimators computed from the two independent samples of sequences* $(\mathbf{S}^1_\ell)_{\ell=1,\dots,n_1}$ *and* $(\mathbf{S}^2_\ell)_{\ell=1,\dots,n_2}$*. We have*

$$\sqrt{\frac{n_1 n_2}{n_1+n_2}}\begin{pmatrix}\widehat{\boldsymbol{p}}^{(1)} - \boldsymbol{p}^1\\ \widehat{\boldsymbol{\theta}}^{(1)} - \boldsymbol{\theta}^1\end{pmatrix} \rightsquigarrow \mathcal{N}\left(0, f\begin{pmatrix}\boldsymbol{\Gamma}^{M_1}_{p^1} & 0\\ 0 & \boldsymbol{\Gamma}^{M_1}_{\theta^1}\end{pmatrix}\right)$$

and

$$\sqrt{\frac{n_1 n_2}{n_1+n_2}}\begin{pmatrix}\widehat{\boldsymbol{p}}^{(2)} - \boldsymbol{p}^2\\ \widehat{\boldsymbol{\theta}}^{(2)} - \boldsymbol{\theta}^2\end{pmatrix} \rightsquigarrow \mathcal{N}\left(0, (1-f)\begin{pmatrix}\boldsymbol{\Gamma}^{M_2}_{p^2} & 0\\ 0 & \boldsymbol{\Gamma}^{M_2}_{\theta^2}\end{pmatrix}\right)$$

and the announced result by taking the difference and the independence assumption of the two samples. □

Proof. of Theorem 4.2. *The proof of Theorem 4.2 is immediate with Lemma 4.1 and the fact that with the continuous mapping theorem* $\widehat{\boldsymbol{\Gamma}}^{f_n}_{p,\theta}$ *converges in probability to* $\boldsymbol{\Gamma}^f_{p,\theta}$*. We deduce that* W_{n_1,n_2} *converges in distribution to a* χ^2 *law with degrees of freedom equal to the rank of* $\boldsymbol{\Gamma}^f_{p,\theta}$*.* □

Proof. of Theorem 4.3.

As in (7), we use the additive structure of the log-likelihood of the sequences $(\mathbf{S}^1_\ell)_{\ell=1,\dots,n_1}$ *and* $(\mathbf{S}^2_\ell)_{\ell=1,\dots,n_2}$ *and decompose the global log-likelihood under the alternative hypothesis into four terms:*

$$\begin{aligned}
-2\log(\lambda_{n_1,n_2}) &= 2\left[\left(n_1\widehat{Q}_{\boldsymbol{p}_1}\left(\widehat{\boldsymbol{p}}^1\right)+n_2\widehat{Q}_{\boldsymbol{p}_2}\left(\widehat{\boldsymbol{p}}^2\right)\right)-\left(n_1\widehat{Q}_{\boldsymbol{p}_1}\left(\boldsymbol{p}^1\right)+n_2\widehat{Q}_{\boldsymbol{p}_2}\left(\boldsymbol{p}^2\right)\right)\right]\\
&\quad+2\left[\left(n_1\widehat{Q}_{\boldsymbol{\theta}_1}\left(\widehat{\boldsymbol{\theta}}^1\right)+n_2\widehat{Q}_{\boldsymbol{\theta}_2}\left(\widehat{\boldsymbol{\theta}}^2\right)\right)-\left(n_1\widehat{Q}_{\boldsymbol{\theta}_1}\left(\boldsymbol{\theta}^1\right)+n_2\widehat{Q}_{\boldsymbol{\theta}_2}\left(\boldsymbol{\theta}^2\right)\right)\right]\\
&\quad-2\left[\left(n_1\widehat{Q}_{\boldsymbol{p}_1}\left(\widehat{\boldsymbol{p}}\right)+n_2\widehat{Q}_{\boldsymbol{p}_2}\left(\widehat{\boldsymbol{p}}\right)\right)-\left(n_1\widehat{Q}_{\boldsymbol{p}_1}\left(\boldsymbol{p}^1\right)+n_2\widehat{Q}_{\boldsymbol{p}_2}\left(\boldsymbol{p}^2\right)\right)\right]\\
&\quad-2\left[\left(n_1\widehat{Q}_{\boldsymbol{\theta}_1}\left(\widehat{\boldsymbol{\theta}}\right)+n_2\widehat{Q}_{\boldsymbol{\theta}_2}\left(\widehat{\boldsymbol{\theta}}\right)\right)-\left(n_1\widehat{Q}_{\boldsymbol{\theta}_1}\left(\boldsymbol{\theta}^1\right)+n_2\widehat{Q}_{\boldsymbol{\theta}_2}\left(\boldsymbol{\theta}^2\right)\right)\right]\\
&= A_p+A_\theta-B_p-B_\theta
\end{aligned}$$

with $A_p = 2\left[\left(n_1\widehat{Q}_{\boldsymbol{p}_1}\left(\widehat{\boldsymbol{p}}^1\right)+n_2\widehat{Q}_{\boldsymbol{p}_2}\left(\widehat{\boldsymbol{p}}^2\right)\right)-\left(n_1\widehat{Q}_{\boldsymbol{p}_1}\left(\boldsymbol{p}^1\right)+n_2\widehat{Q}_{\boldsymbol{p}_2}\left(\boldsymbol{p}^2\right)\right)\right]$,
$A_\theta = 2\left[\left(n_1\widehat{Q}_{\boldsymbol{\theta}_1}\left(\widehat{\boldsymbol{\theta}}^1\right)+n_2\widehat{Q}_{\boldsymbol{\theta}_2}\left(\widehat{\boldsymbol{\theta}}^2\right)\right)-\left(n_1\widehat{Q}_{\boldsymbol{\theta}_1}\left(\boldsymbol{\theta}^1\right)+n_2\widehat{Q}_{\boldsymbol{\theta}_2}\left(\boldsymbol{\theta}^2\right)\right)\right]$,
$B_p = 2\left[\left(n_1\widehat{Q}_{\boldsymbol{p}_1}\left(\widehat{\boldsymbol{p}}\right)+n_2\widehat{Q}_{\boldsymbol{p}_2}\left(\widehat{\boldsymbol{p}}\right)\right)-\left(n_1\widehat{Q}_{\boldsymbol{p}_1}\left(\boldsymbol{p}^1\right)+n_2\widehat{Q}_{\boldsymbol{p}_2}\left(\boldsymbol{p}^2\right)\right)\right]$ *and*
$B_\theta = 2\left[\left(n_1\widehat{Q}_{\boldsymbol{\theta}_1}\left(\widehat{\boldsymbol{\theta}}\right)+n_2\widehat{Q}_{\boldsymbol{\theta}_2}\left(\widehat{\boldsymbol{\theta}}\right)\right)-\left(n_1\widehat{Q}_{\boldsymbol{\theta}_1}\left(\boldsymbol{\theta}^1\right)+n_2\widehat{Q}_{\boldsymbol{\theta}_2}\left(\boldsymbol{\theta}^2\right)\right)\right]$.

We can note that the estimators $(\widehat{\mathbf{p}}^1,\widehat{\boldsymbol{\theta}}^1)$ *and* $(\widehat{\mathbf{p}}^2,\widehat{\boldsymbol{\theta}}^2)$ *satisfy the first order conditions* $\nabla_{\boldsymbol{p}_1}\widehat{Q}_{\boldsymbol{p}_1}(\widehat{\boldsymbol{p}}^1)=0$ *and* $\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\widehat{\boldsymbol{\theta}}^1)=0$ *and* $\nabla_{\boldsymbol{p}_2}\widehat{Q}_{\boldsymbol{p}_2}(\widehat{\boldsymbol{p}}^2)=0$ *and* $\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\widehat{\boldsymbol{\theta}}^2)=0$ *whereas the estimators* $(\widehat{\mathbf{p}},\widehat{\boldsymbol{\theta}})$ *satisfy* $n_1\nabla_{\boldsymbol{p}_1}\widehat{Q}_{\boldsymbol{p}_1}(\widehat{\boldsymbol{p}})+n_2\nabla_{\boldsymbol{p}_2}\widehat{Q}_{\boldsymbol{p}_2}(\widehat{\boldsymbol{p}})=0$ *and* $n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\widehat{\boldsymbol{\theta}})+n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\widehat{\boldsymbol{\theta}})=0$.

We present only the proof corresponding to the part of the likelihood ratio related to the sojourn times. The same calculus would lead to similar results for the transition matrices and are not presented.

Study of A_θ**.** *Using a Taylor expansion of the log-likelihood around* $\widehat{\boldsymbol{\theta}}^1$ *and* $\widehat{\boldsymbol{\theta}}^2$*, we get, similarly to (31),*

$$A_\theta = \begin{pmatrix}\boldsymbol{\theta}^1-\widehat{\boldsymbol{\theta}}^1\\ \boldsymbol{\theta}^2-\widehat{\boldsymbol{\theta}}^2\end{pmatrix}^\top\left[-\begin{pmatrix}n_1\nabla_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\overline{\boldsymbol{\theta}}^1) & 0\\ 0 & n_2\nabla_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\overline{\boldsymbol{\theta}}^2)\end{pmatrix}\right]\begin{pmatrix}\boldsymbol{\theta}^1-\widehat{\boldsymbol{\theta}}^1\\ \boldsymbol{\theta}^2-\widehat{\boldsymbol{\theta}}^2\end{pmatrix}.$$

To find the asymptotic distribution of $\begin{pmatrix}\boldsymbol{\theta}^1-\widehat{\boldsymbol{\theta}}^1\\ \boldsymbol{\theta}^2-\widehat{\boldsymbol{\theta}}^2\end{pmatrix}^\top$ *we make a Taylor expansion about* $(\widehat{\boldsymbol{\theta}}^1,\widehat{\boldsymbol{\theta}}^2)$ *of the gradient of the log-likelihood evaluated at* $(\boldsymbol{\theta}^1,\boldsymbol{\theta}^2)$*. We have, for parameters* $\boldsymbol{\theta}_1$ *and* $\boldsymbol{\theta}_2$*,*

$$\begin{pmatrix}n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}^{\mathbf{1}})\\ n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta}^{\mathbf{2}})\end{pmatrix}=0+\begin{pmatrix}n_1\nabla_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\overline{\boldsymbol{\theta}}^1) & 0\\ 0 & n_2\nabla_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\overline{\boldsymbol{\theta}}^2)\end{pmatrix}\begin{pmatrix}\boldsymbol{\theta}^1-\widehat{\boldsymbol{\theta}}^1\\ \boldsymbol{\theta}^2-\widehat{\boldsymbol{\theta}}^2\end{pmatrix}\tag{33}$$

which enables us to obtain

$$A_\theta = -\frac{1}{n_1+n_2}\begin{pmatrix} n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1}) \\ n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2}) \end{pmatrix}^\top \begin{pmatrix} \frac{n_1+n_2}{n_1}\nabla_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\overline{\boldsymbol{\theta}}^1) & 0 \\ 0 & \frac{n_1+n_2}{n_2}\nabla_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\overline{\boldsymbol{\theta}}^2) \end{pmatrix} \begin{pmatrix} n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1}) \\ n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2}) \end{pmatrix}.$$

Thanks to the weak law of large numbers and the continuous mapping Theorem, we have the convergence in probability $-\nabla_{\boldsymbol{\theta}_1\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\overline{\boldsymbol{\theta}_1}) \to \left(\boldsymbol{\Gamma}_{\theta_1}^{M_1}\right)^{-1}$ *and* $-\nabla_{\boldsymbol{\theta}_2\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\overline{\boldsymbol{\theta}_2}) \to \left(\boldsymbol{\Gamma}_{\theta_2}^{M_2}\right)^{-1}$. *We have from (29) in the proof of Theorem 3.2, that* $\begin{pmatrix} \frac{n_1}{\sqrt{n_1+n_2}}\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1}) \\ \frac{n_2}{\sqrt{n_1+n_2}}\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2}) \end{pmatrix}$ *is bounded in probability. We obtain finally*

$$A_\theta = \frac{1}{n_1+n_2}\begin{pmatrix} n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1}) \\ n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2}) \end{pmatrix}^\top \boldsymbol{\Gamma}_{\theta_1,\theta_2}^f \begin{pmatrix} n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1}) \\ n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2}) \end{pmatrix} + o_p(1)$$

with $\boldsymbol{\Gamma}_{\theta_1,\theta_2}^f = \begin{pmatrix} \frac{1}{1-f}\boldsymbol{\Gamma}_{\theta_1}^{M_1} & 0 \\ 0 & \frac{1}{f}\boldsymbol{\Gamma}_{\theta_2}^{M_2} \end{pmatrix}$.

Study of B_θ. *The hypothesis H_0 can be written using a differentiable map of dimension r, with $H_0 : a\left(\boldsymbol{\theta^1}, \boldsymbol{\theta^2}\right) = 0$. We suppose that 0 is a regular value of the map, thus we can find an injective mapping ϕ, which is $\mathcal{C}^2$, such as $a(\phi(x)) = 0$ and $\boldsymbol{x^0}$ is the only value that checks $\phi(\boldsymbol{x^0}) = \left(\boldsymbol{\theta^1}, \boldsymbol{\theta^2}\right)$ of dimension $2k-r$ with k the dimension of $\boldsymbol{\theta^1}$ which is the same as the one of $\boldsymbol{\theta^2}$. Thus, $k = D(D-1)d$ when there is no absorbing state and $k = (D-1)^2 d$ else.*

We set $\widehat{Q}(\boldsymbol{\theta^1}, \boldsymbol{\theta^2}) = n_1\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1}) + n_2\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2})$. *Then, B_θ can be written as:*

$$\begin{aligned} B_\theta &= 2[\widehat{Q}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}}) - \widehat{Q}(\boldsymbol{\theta^1}, \boldsymbol{\theta^2})] \\ &= 2[\widehat{Q}(\phi(\widehat{\mathbf{x}})) - \widehat{Q}(\phi(\boldsymbol{x^0}))] \end{aligned}$$

where $\phi(\widehat{\mathbf{x}}) = (\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\theta}})$ *and* $\widehat{\mathbf{x}}$ *is a consistent estimator of* $\boldsymbol{x^0}$. *We can note that the gradient and the hessian matrix are:*

$$\begin{aligned} \nabla_{\boldsymbol{x}}\left(\widehat{Q}\circ\phi\right)(\boldsymbol{x}) &= \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})\nabla_{\phi(\boldsymbol{x})}\widehat{Q}(\phi(\boldsymbol{x})) \\ \nabla_{\boldsymbol{x}\boldsymbol{x}}\left(\widehat{Q}\circ\phi(\boldsymbol{x})\right) &= \nabla_{\boldsymbol{x}\boldsymbol{x}}\left(\phi(\boldsymbol{x})\right)\nabla_{\phi(\boldsymbol{x})}\widehat{Q}(\phi(\boldsymbol{x})) + \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})\nabla_{\phi(\boldsymbol{x})\phi(\boldsymbol{x})}\widehat{Q}(\phi(\boldsymbol{x}))\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})^\top. \end{aligned}$$

Thanks to a Taylor expansion of $\widehat{Q}(\phi(x))$, *we obtain,*

$$B_\theta = (\boldsymbol{x^0} - \widehat{\mathbf{x}})^\top\left[-\nabla_{\boldsymbol{x}\boldsymbol{x}}\left(\widehat{Q}\circ\phi(\overline{\mathbf{x}})\right)\right](\boldsymbol{x^0} - \widehat{\mathbf{x}})$$

where $\overline{\mathbf{x}}$ *belongs elementwise to segment defined by* $\boldsymbol{x^0}$ *and* $\widehat{\mathbf{x}}$. *We make a Taylor expansion of* $\nabla_{\boldsymbol{x}}\left(\widehat{Q}\circ\phi\right)$ *in order to obtain the law of* $\boldsymbol{x^0} - \widehat{\mathbf{x}}$ *:*

$$\nabla_{\boldsymbol{x}}\left(\widehat{Q}\circ\phi\right)(\boldsymbol{x^0})=\nabla_{\boldsymbol{x}}\left(\widehat{Q}\circ\phi\right)(\widehat{\mathbf{x}})+\nabla_{\boldsymbol{xx}}\left(\widehat{Q}\circ\phi\right)(\overline{\mathbf{x}})(\boldsymbol{x^0}-\widehat{\mathbf{x}}).$$

Then, we have

$$(\boldsymbol{x^0}-\widehat{\mathbf{x}})^\top=\left[\nabla_{\boldsymbol{x}}\left(\widehat{Q}\circ\phi\right)(\boldsymbol{x^0})\right]^\top\left[\nabla_{\boldsymbol{xx}}\left(\widehat{Q}\circ\phi\right)(\overline{\mathbf{x}})\right]^{-1}$$

and

$$\begin{aligned}B_\theta&=\left[\nabla_{\boldsymbol{x}}\left(\widehat{Q}\circ\phi\right)(\boldsymbol{x^0})\right]^\top\left[-\nabla_{\boldsymbol{xx}}\left(\widehat{Q}\circ\phi\right)(\overline{\mathbf{x}})\right]^{-1}\nabla_{\boldsymbol{x}}\left(\widehat{Q}\circ\phi\right)(\boldsymbol{x^0})\\&=\left[\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})\begin{pmatrix}n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1})\\n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2})\end{pmatrix}\right]^\top\left[-\nabla_{\boldsymbol{xx}}\left(\widehat{Q}\circ\phi\right)(\overline{\mathbf{x}})\right]^{-1}\left[\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})\begin{pmatrix}n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1})\\n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2})\end{pmatrix}\right].\end{aligned}$$

As $\overline{\mathbf{x}}$ is consistent and thanks to the continuous mapping theorem, we get the convergence in probability,

$$\frac{1}{n_1+n_2}\left[-\nabla_{\boldsymbol{xx}}\left(\widehat{Q}\circ\phi\right)(\overline{\mathbf{x}})\right]\to\left[\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})\right]^\top\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0}).$$

Finally, we have,

$$\begin{aligned}B_\theta=&\frac{1}{n_1+n_2}\begin{pmatrix}n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1})\\n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2})\end{pmatrix}^\top\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})\left[\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})^\top\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})\right]^{-1}\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})^\top\begin{pmatrix}n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1})\\n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2})\end{pmatrix}\\&+o_p(1).\end{aligned}$$

Study of the test statistics. *We deduce from (29), the convergence in distribution*

$$\frac{1}{\sqrt{n_1+n_2}}\left(\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\right)^{\frac{1}{2}}\begin{pmatrix}n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1})\\n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2})\end{pmatrix}\rightsquigarrow\mathcal{N}\left(0,\mathbf{I}_{2k\times 2k}\right).$$

Thus, the test statistics (considering only the sojourn times) can be written as follows,

$$\begin{aligned}A_\theta-B_\theta=&\frac{1}{\sqrt{n_1+n_2}}\begin{pmatrix}n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1})\\n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2})\end{pmatrix}^\top\left(\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\right)^{\frac{1}{2}}\boldsymbol{M}_\theta\left(\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\right)^{\frac{1}{2}}\begin{pmatrix}n_1\nabla_{\boldsymbol{\theta}_1}\widehat{Q}_{\boldsymbol{\theta}_1}(\boldsymbol{\theta^1})\\n_2\nabla_{\boldsymbol{\theta}_2}\widehat{Q}_{\boldsymbol{\theta}_2}(\boldsymbol{\theta^2})\end{pmatrix}\frac{1}{\sqrt{n_1+n_2}}\\&+o_p(1)\end{aligned}$$

with $\mathbf{M}_\theta=\boldsymbol{I}_{2k\times 2k}-\left(\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\right)^{-\frac{1}{2}}\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})\left[\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})^\top\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})\right]^{-1}\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})^\top\left(\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\right)^{-\frac{1}{2}}$.

Noting that $\mathbf{M}_\theta$ is an orthogonal projection matrix, the asymptotic distribution of $A_\theta-B_\theta$ is a χ^2 law whose degrees of freedom is equal to the rank of $\mathbf{M}_\theta$, which is equal to its

trace,

$$\begin{aligned}
Tr(\mathbf{M}_\theta) &= Tr\left(\boldsymbol{I}_{2k\times 2k} - \left(\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\right)^{-\frac{1}{2}} \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0}) \left[\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})^\top \boldsymbol{\Gamma}^f_{\theta_1,\theta_2} \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})\right]^{-1} \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})^\top \left(\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\right)^{-\frac{1}{2}}\right) \\
&= Tr(\boldsymbol{I}_{2k\times 2k}) - Tr\left(\left(\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\right)^{-\frac{1}{2}} \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0}) \left[\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})^\top \boldsymbol{\Gamma}^f_{\theta_1,\theta_2} \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})\right]^{-1} \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})^\top \left(\boldsymbol{\Gamma}^f_{\theta_1,\theta_2}\right)^{-\frac{1}{2}}\right) \\
&= Tr(\boldsymbol{I}_{2k\times 2k}) - Tr\left(\left[\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})^\top \boldsymbol{\Gamma}^f_{\theta_1,\theta_2} \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})\right] \left[\nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})^\top \boldsymbol{\Gamma}^f_{\theta_1,\theta_2} \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x^0})\right]^{-1}\right) \\
&= r.
\end{aligned}$$

Similarly to the case of the sojourn time parameters and the case of the transition probabilities for one sample, we obtain for the part of the transition probabilities a matrix $\mathbf{M}_p$ *whose rank is equal to* $D(D-2)$ *when there is no absorbing state and* $(D-1)(D-2)$ *else.*

In conclusion, we have, thanks to the asymptotic normality and the block diagonal structure of the asymptotic variance that the test statistics has, under the null hypothesis, a χ^2 *distribution whose degrees of freedom is the sum of the ranks of* $\mathbf{M}_\theta$ *and of* $\mathbf{M}_p$*.*

□