JOURNAL OF APPLIED STATISTICS
2025, VOL. 52, NO. 11, 2172–2190
https://doi.org/10.1080/02664763.2025.2457008

Taylor & Francis
Taylor & Francis Group

Gradient test to assess homogeneity of probabilities in discrete-time transition models with application in agricultural science data

Laura Vicuña Torres de Paula[a*], Idemauro Antonio Rodrigues de Lara[a*], Cesar Auguto Taconeli[b], Carolina Reigada[c] and Rafael de Andrade Moral[d]

[a]"Luiz de Queiroz" College, University of São Paulo, Piracicaba, Brazil;; [b]Statistics Department, Federal University of Paraná, Curitiba, Brazil; [c]Ecology and Evolutionary Biology Department, Federal University of São Carlos, São Carlos, Brazil; [d]Department of Mathematics and Statistics, Maynooth University, Maynooth, Ireland

ABSTRACT

Longitudinal studies in discrete or continuous time involving categorical data are common in agricultural sciences. Transition models can be used as a means to analyse the resulting data, especially when the aim is to describe category changes over time, as well as to accommodate covariates due to experimental design. Here we focus on discrete-time models, for which it is critical to assess whether the underlying process is stationary or not. Tests based on likelihood procedures are very useful, and here we propose the Gradient test to assess stationary, or homogeneity of transition probabilities. We carried out simulation studies to evaluate the performance of the proposed test, which indicated a good performance regarding type-I error and power when compared to other classical tests available in the literature. As motivation we present two studies with agricultural data, the first one applied to entomology with nominal responses and the second application refers to the degree of injury in pigs. Using our proposed test, stationarity and non-stationarity were verified respectively in the applications. Since the gradient test to assess stationarity has a simplified structure when compared to other tests, it is therefore a useful alternative when carrying out inference in these types of models.

ARTICLE HISTORY
Received 23 June 2023
Accepted 11 January 2025

KEYWORDS
Longitudinal categorical data; maximum likelihood estimation; stochastic processes; simulation studies; biological control

MATHEMATIC SUBJECT CLASSIFICATIONS
92-05; 92-08; 62M02

1. Introduction

Categorical data are common in experimental studies, especially in agricultural sciences, where behavioural studies are carried out using animals and insects, associated with welfare practices or biological pest control. The design of these studies is usually longitudinal over time, in order to observe the effectiveness of treatments and effects of the associated covariates [7].

CONTACT Idemauro Antonio Rodrigues de Lara idemauro@usp.br "Luiz de Queiroz" College, University of São Paulo, Piracicaba, Brazil

*These authors contributed equally to this work.

In longitudinal studies, it is necessary to consider the possible dependence between repeated measurements over time [6,22]. There are many approaches that can be used to this end; for instance, hidden Markov models with covariates represent a flexible framework within the wide class of Markov-type models [4]. Among this class of models, transition models can be used to describe the changes of response categories over time as well as the influence of covariates [22]. According to Ref. [14], the transition models are based on stochastic processes and generalized linear models, by simultaneously considering models for discrete variables and the Markovian property in the linear predictor. Indeed, for nominal responses, the generalized logit model is the appropriate choice, whereas when the response is ordinal, cumulative models [21] can be employed. Ware *et al.* [31] is a classical reference in this context, in which the proportional odds regression model (a particular case of cumulative logit models) is considered with the inclusion of the subject's response at the previous time as an additional covariate in the linear predictor (first-order Markov chain).

Although they are very common in applications in health sciences (e.g. see Refs [2,9,11,18,32,33]), transition models have been applied to several other fields, including agricultural sciences. For instance, the authors [14] employed the proportional odds transition model to assess the lesion condition of pigs, exposed or not to environmental enrichment, and the authors in Refs [16,23] applied the generalized logit transition model to understand the movement patterns of female adults of *Diaphorina citri*, a pest of citrus plantations.

Despite their importance, there are methodological questions that should be studied to improve the applicability of transition models. One of them is the condition of stationarity or homogeneity of the transition probabilities over time. In general, this assumption is made to simplify the model fitting process and to reduce the number of parameters. However, this is a strong assumption that can change the conclusions. For instance, under stationarity, we may erroneously conclude that the effect of a treatment given the previous response is persistent over time, when in fact it loses its effect. Therefore, it is imperative to evaluate the stationarity condition in transition models. For discrete-time models this condition is equivalent to the homogeneity of transition probability matrices.

In the literature, there are tests to assess stationarity, such as the classical test proposed by Ref. [3], based on the estimated transition probabilities. The logic of this test is that under stationarity the transition probabilities over time are homogeneous, in such a way that the $T-1$ transition probabilities matrices (under a first-order Markov property and T time occasions) can be summarized by a single matrix. More recently, the authors [15] proposed Wald-type and likelihood-ratio tests based on regression coefficients of generalized transition model, that alternatively to the classical one, do not require the computation of the estimated probabilities.

The aim of this article is to present the Gradient test to verify stationarity in discrete-time transition models, given that such kind of a test has not yet been developed for this purpose. The Gradient statistic was originally proposed by Ref. [29] and is a composition of the Score [24] and modified Wald [10] tests, resulting in a simplified expression, with attractive asymptotic properties for practical applications [20]. Therefore, the Gradient statistic is an alternative to the Score, Likelihood Ratio and Wald tests. We present a brief review of the Gradient statistic and its properties in Section 3. The remainder of this paper is structured as follows: in Section 2, we present a review of the discrete-time stochastic process, transition models for nominal and ordinal data, and hypothesis tests for stationarity; in Section 3, we present the proposed gradient test; simulation studies and comparison with other available tests are presented in Section 4; in Sections 5.1 and 5.2, we present two case studies in entomology and animal science, respectively, both in the context of agricultural sciences, as well as results and discussion. Finally, in section 6, we make final considerations.

2. A review on transition models

2.1. Discrete-time stochastic process and Markov chains

A stochastic process is described as a sequence of events governed by probabilistic laws over discrete or continuous time [13]. Briefly, according to Ref. [27], it can be defined as a collection of random variables indexed by time, i.e. Y_t, where $t \in \tau$. When τ is enumerable, it is a discrete-time process, otherwise a continuous-time process. For our purposes, we consider a discrete variable and time process, that is, Y_t takes values in state space $S = \{1, 2, \ldots, k\}$, that represent the response categories of the study, and $\tau = \{1, \ldots, T-1, T\}$ the specific observation times. We also assume a first-order Markov process, defined by:

$$\pi_{ab}(t-1,t) = P(Y_t = b \mid Y_1 = y_1, Y_2 = y_2, \ldots, Y_{(t-1)} = a) = P(Y_t = b \mid Y_{(t-1)} = a),$$

where $\pi_{ab}(t-1,t) \forall\, a, b \in S$, are the transition probabilities. To simplify the notation, the probabilities can be rewritten as $\pi_{ab}(t-1,t) = \pi_{ab}(t)$ and in matrix form

$$\boldsymbol{P}(t) = \begin{pmatrix} \pi_{11}(t) & \pi_{12}(t) & \ldots & \pi_{1k}(t) \\ \pi_{21}(t) & \pi_{22}(t) & \ldots & \pi_{2k}(t) \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{k1}(t) & \pi_{k2}(t) & \ldots & \pi_{kk}(t) \end{pmatrix}$$

where each element corresponds to a transition and satisfy, $\pi_{ab}(t) \geq 0, a, b = 1, 2, \ldots, k$ and $\sum_{b=1}^{k} \pi_{ab}(t) = 1$. Then, these probabilities may or may not be homogeneous over time, characterizing stationary and non-stationary processes. When the probability matrices are homogeneous over time, it is denoted by $\boldsymbol{P}$, with elements π_{ab}, where the absence of the term (t) indicates independence with respect to time [15].

2.2. Discrete-time transition models

When we have the experimental design structure or the presence of covariates, the best way to implement a transition model is through generalized linear models (GLMs) and extensions. The Markov property is operationalized by including the previous time response (or some function of it) as an additional covariate in the linear predictor [31]. Hence, if $\boldsymbol{y}_{it} = (y_{i1}, y_{i2}, \ldots, y_{in_i})^T$ is an individual profile vector of dimension $(n_i \times 1)$, $\boldsymbol{x}_{it} = (x_{it1}, x_{it2}, \ldots, x_{itp})^T$ is a $(p \times 1)$ vector of covariates, and $\boldsymbol{h}_{it} = (y_{i(t-1)}, y_{i(t-2)}, \ldots, y_{i(t-q)})$ is the individual history, the transition model is defined for the random variable $Y_{it} \mid \boldsymbol{h}_{it}$ as [6]

$$\eta_{it} = g(\mu_{it}^C) = \mathbf{x}_{it}^T \boldsymbol{\beta} + \sum_{r=1}^{s} f_r^*(\mathbf{h}_{it}; \boldsymbol{\alpha})$$

where $g(\mu_{it}^C) = g(\mathrm{E}(Y_{it}|\mathbf{h}_{it}))$ represents an appropriate link function, $f_r^*(\boldsymbol{h}_{it}; \boldsymbol{\alpha})$ is a linear function of previous responses, which may also include interactions with the covariates. In the simplest case $f_r^*(\boldsymbol{h}_{it}; \boldsymbol{\alpha}) = \alpha_r f_r^*(\boldsymbol{h}_{it}) = \alpha y_{(t-1)}$. Therefore, the parameters of interest are denoted by $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\alpha})$. Similarly to GLMs for independent data, the authors [6,22] describe in detail the estimation process for the stationary case, using a stacked dataset structure, with an additional column for the previous response. For the non-stationary case, the authors [31] suggest fitting a model for each time occasion, always considering the previous response as an additional covariate, and the authors [26] present several ways to define the linear predictor for stationary or non-stationary processes.

For polytomous categorical variables (more than two response categories) we can extend the generalized logit models [1] and include in the linear predictor some function of the previous response to compose the transition model. Considering the initial notation already established but now, with $Y_{it} \in S = \{1, 2, \ldots, k\}$ and a first-order Markovian property, in its simplest form, we define the models by distinguishing between nominal and ordinal cases.

Transition model for nominal responses

In this case, the generalized logits transition model is defined by the equation:

$$\eta = \log\left(\frac{\pi_{ab}(t)}{\pi_{ak}(t)}\right) = \lambda_{b(t)} + \boldsymbol{\beta}_{b(t)}^T \boldsymbol{x} + \alpha_{b(t)} y_{(t-1)} \quad (b = 1, \ldots, k-1), \tag{1}$$

that represents the transition given a previous state a to state b using category k as reference, where $a, b = 1, \ldots, k$, $\boldsymbol{\lambda}_{(t)} = (\lambda_{1t}, \lambda_{2t}, \ldots, \lambda_{(k-1)t})^T$ is the vector of intercepts for each category (called the perturbation parameter in categorical models), $\boldsymbol{\beta}_{b(t)} = (\beta_{b1t}, \beta_{b2t}, \ldots, \beta_{bpt})^T$ is the vector of unknown parameters associated with covariates, and $\boldsymbol{\alpha}_{(t)} = (\alpha_{1t}, \alpha_{2t}, \ldots, \alpha_{(k-1)t})^T$ is the vector of parameters related to the Markov covariate (lagged response). To simplify the notation, consider $\boldsymbol{\delta}_{b(t)} = (\boldsymbol{\beta}_{b(t)}, \boldsymbol{\alpha}_{(t)})^T$ the vector of unknown parameters associated with $\boldsymbol{x}^* = (\boldsymbol{x}, y_{(t-1)})$ (vector of covariates) and $\boldsymbol{\theta}_{b(t)} = (\boldsymbol{\lambda}_{(t)}, \boldsymbol{\delta}_{b(t)})^T$ the vector of unknown parameters that depend of b, including the perturbation ones. Thus, once $\boldsymbol{\theta}_{b(t)}$ is estimated for each transition occasion, $t = 1, 2, \ldots, T-1$, the transition probabilities are computed via:

$$\hat{\pi}_{ab}(t) = \frac{\exp(\hat{\lambda}_{b(t)} + \hat{\boldsymbol{\delta}}_{b(t)}\boldsymbol{x}^*)}{1 + \sum_{b=1}^{k-1} \exp(\hat{\lambda}_{b(t)} + \hat{\boldsymbol{\delta}}_{b(t)}\boldsymbol{x}^*)}, \quad b = 1, 2, \ldots, k-1$$

and for $b = k$, we have:

$$\hat{\pi}_{ak}(t) = \frac{1}{1 + \sum_{b=1}^{k-1} \exp(\hat{\lambda}_{b(t)} + \hat{\boldsymbol{\delta}}_{b(t)}\boldsymbol{x}^*)},$$

where $\hat{\lambda}_{b(t)}$ and $\hat{\boldsymbol{\delta}}_{b(t)}$ are the estimates of the parameters for which $\hat{\pi}_{ab}(t)$ are estimated for all previous states $a \in S$.

Transition model for ordinal responses

For ordinal responses, we use cumulative logit models, the simplest being the proportional-odds regression model. Let $\gamma_{ab}(t) = \pi_{a1}(t) + \cdots + \pi_{ab}(t)$ be the cumulative probability until category $b \in S$ from a state, with $1 < 2 < \cdots < k$, the proportional-odds transition model is given by:

$$\eta = \log\left(\frac{\gamma_{ab}(t)}{1-\gamma_{ab}(t)}\right) = \lambda_{b(t)} + \boldsymbol{\beta}_{(t)}^{T}\boldsymbol{x} + \alpha_{(t)}y_{(t-1)} \quad (b = 1,\ldots,k-1), \tag{2}$$

where now only the intercept $\lambda_{b(t)}$ varies per each response category. Just as in the nominal case, we can use the vector notation $\boldsymbol{\delta}_{(t)} = (\boldsymbol{\beta}_{(t)}, \alpha_{(t)})^T$ for the parameters associated with the covariates including the previous response (Markov), and $\boldsymbol{\theta}_{b(t)} = (\boldsymbol{\lambda}_{(t)}, \boldsymbol{\delta}_{(t)})^T$ to represent the vector of unknown parameters linked to the intercepts $\lambda_{b(t)}, b = 1, 2, \ldots, k-1$. From model 2, the estimated cumulative transition probabilities are:

$$\hat{\gamma}_{ab}(t) = \frac{\exp(\hat{\lambda}_{b(t)} + \hat{\boldsymbol{\delta}}_t^T\boldsymbol{x}^*)}{1+\exp(\hat{\lambda}_{b(t)} + \hat{\boldsymbol{\delta}}_t^T\boldsymbol{x}^*)}, \quad b = 1, 2, \ldots, k-1$$

where $\hat{\lambda}_{b(t)}$ and $\hat{\boldsymbol{\delta}}_{(t)}$ are the parameter estimates. Then, the transition probabilities for each category (for $a \in S$), are given by $\hat{\pi}_{ab}(t) = \hat{\gamma}_{ab}(t) - \hat{\gamma}_{a(b-1)}(t)$, and in particular for $b = k$ we have: $\hat{\pi}_{ak}(t) = 1 - \hat{\gamma}_{a(k-1)}(t)$. Models (1) and (2) can be fitted by maximum likelihood via iterative processes, as per usual in models for categorical data [12].

2.3. Tests to assess stationarity

We begin by stating the assumptions regarding the stationarity condition, in terms of homogeneity of transition probabilities. The hypotheses are defined as follows:

$$\begin{aligned} &\mathrm{H}_0 : \pi_{ab}(t) = \pi_{ab}(0), \quad \text{for all } t = (1,\ldots,T), \text{ and } a, b, \in (1,\ldots,k);\\ &\mathrm{H}_1 : \pi_{ab}(t) \neq \pi_{ab}(l), \quad \text{for some } t \neq l, \end{aligned} \tag{3}$$

where $\pi_{ab}(0)$ represents the homogeneous transition probabilities over time, or in matrix notation:

$$\begin{aligned} &\mathrm{H}_0 : \boldsymbol{P}(t) = \boldsymbol{P}(0), \quad \text{for all } t = 1,\ldots,T-1;\\ &\mathrm{H}_1 : \boldsymbol{P}(t) \neq \boldsymbol{P}(l), \quad \text{for some } (t,l) \in (1,\ldots,T-1) \quad \text{where } t \neq l. \end{aligned}$$

The classical test statistic proposed by [3] for these hypotheses (3) is:

$$\xi = \sum_{t=1}^{T}\sum_{a=1}^{k}\sum_{b=1}^{k} \frac{n_a(t-1)[\hat{\pi}_{ab}(t) - \hat{\pi}_{ab}(0)]^2}{\hat{\pi}_{ab}(0)}, \tag{4}$$

where $n_a(t-1)$ is the number of individuals who were in state a in time $(t-1)$, $\hat{\pi}_{ab}(t)$ are the estimated probabilities for each time transition and $\hat{\pi}_{ab}$ represents the transition probability under H_0. According to Ref. [3], for nominal data, the statistic ξ has an asymptotic χ^2_{υ} distribution with $\upsilon = k(k-1)(T-1)$. For ordinal data, the number of degrees of freedom may differ due to the cutoff points, as discussed by Ref. [15].

As alternatives to the classic (4), the authors [15] have recently proposed tests that do not require estimates of the transition probabilities. They are based directly on the estimated parameters $\boldsymbol{\theta}_t$ of the transition model. For these tests, the hypotheses are rewritten as

$$\begin{aligned} &\mathrm{H}_0 : \boldsymbol{\theta}_t = \boldsymbol{\theta}_0, \quad \text{for all } t = 1, \dots, T-1; \\ &\mathrm{H}_1 : \boldsymbol{\theta}_t \neq \boldsymbol{\theta}_l, \quad \text{for some } t \neq l, \end{aligned} \tag{5}$$

where $\boldsymbol{\theta}_0$ is the vector of parameter values under stationarity. The tests are based on the likelihood theory, using the asymptotic result $\hat{\boldsymbol{\theta}}_t \sim \mathrm{N}(\boldsymbol{\theta}_t, \mathcal{F}_t^{-1})$, where $\mathcal{F}_{\boldsymbol{t}}$ is the Fisher information matrix in the tth transition. The likelihood-ratio (LR) test statistic is given by:

$$\Lambda = -2\left[\log(L(\hat{\boldsymbol{\theta}}_0, \boldsymbol{x})) - \sum_{t=1}^{T-1} \log(L_t(\hat{\boldsymbol{\theta}}_t, \boldsymbol{x}))\right] \tag{6}$$

where $\hat{\boldsymbol{\theta}}_0$ is the estimated parameter vector under H_0 and $\hat{\boldsymbol{\theta}}_t$ is the estimated parameter vector for time transition t.

The Wald test statistic for the tth transition is given by

$$W_t = [\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0]^T \hat{\mathcal{F}}_{\hat{\boldsymbol{\theta}}_t} [\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0], \quad t = 1, 2, \dots, T-1, \tag{7}$$

where $\hat{\mathcal{F}}_{\hat{\boldsymbol{\theta}}_t}$ is the evaluated Fisher information matrix using $\hat{\boldsymbol{\theta}}_t$. Then, the Wald test statistic considering all time transitions is $W = \sum_{t=1}^{T-1} W_t$.

Under regularity conditions, Λ and W are asymptotically distributed as χ^2_υ, with $\upsilon = dim(\Theta) - dim(\Theta_0)$. According to Ref. [15], a problem with the W statistic is that in practice $\boldsymbol{\theta}_0$ is unknown, and an alternative would be to use the estimated value under the null hypothesis, i.e. assume stationarity and use $\hat{\boldsymbol{\theta}}_0$ as the true value $\boldsymbol{\theta}_0$. An alternative way to get around this problem, also demonstrated by Ref. [15] is to consider the null hypothesis in the following form:

$$\mathrm{H}_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}_3 = \cdots = \boldsymbol{\theta}_{T-1}, \tag{8}$$

versus the alternative that at least one pair of parameter vectors are different. In this way, we do not need $\boldsymbol{\theta}_0$ and there are $T-1$ non-redundant comparisons of the first-order transitions, one of which is:

$$W_t = [\hat{\boldsymbol{\theta}}_t - \hat{\boldsymbol{\theta}}_1]^T \hat{\mathcal{F}}_{\hat{\boldsymbol{\theta}}_t} [\hat{\boldsymbol{\theta}}_t - \hat{\boldsymbol{\theta}}_1], \quad t = 2, \dots, T-1,$$

where $\hat{\boldsymbol{\theta}}_1$ corresponds to the estimate in the first transition and the matrix $\hat{\mathcal{F}}_{\hat{\boldsymbol{\theta}}_t}$ has elements given by $\mathrm{Var}(\hat{\boldsymbol{\theta}}_t, \hat{\boldsymbol{\theta}}_{t'}) = \mathrm{Var}(\hat{\boldsymbol{\theta}}_t) + \mathrm{Var}(\hat{\boldsymbol{\theta}}_{t'})$ and $\mathrm{Cov}(\hat{\boldsymbol{\theta}}_{t''} - \hat{\boldsymbol{\theta}}_t, \hat{\boldsymbol{\delta}}_{t''} - \hat{\boldsymbol{\theta}}_{t'}) = \mathrm{Var}(\hat{\boldsymbol{\theta}}_{t''})$, for all $t \neq t' \neq t''$, given that the transitions occur independently. Moreover, the simulation studies developed by Ref. [15] demonstrated the competitive performance of these tests when compared to the classical one. Additionally, the authors [26] present an alternative way of evaluating stationarity from the stacked data structure, using a sequence of nested models, containing interactions between the factors studied and time, thus allowing the identification of sources of variation in the probability matrices. However, the purposes of this work are centred on the presentation and application of the gradient test to identify whether a process is stationary or not.

3. Gradient statistic

In this section, we present a brief introduction to the Gradient Statistics as a review, and later we extend it to evaluate the condition of homogeneity of probabilities in transition models.

Consider a classical hypothesis test, in situations where it is not possible to apply an exact test, such as: $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, for p-dimensional vectors. The modified Wald [10] and Score [24] tests are given, respectively, by:

$$W_M = [\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0]^T \mathcal{F}(\boldsymbol{\theta}_0)[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0], \tag{9}$$

and

$$S = \boldsymbol{U}'(\boldsymbol{\theta}_0)\mathcal{F}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{U}(\boldsymbol{\theta}_0) \tag{10}$$

where the Fisher information matrix, $\mathcal{F}(\boldsymbol{\theta}_0)$, is constructed as a function of the vector tested in the null hypothesis. Let $\boldsymbol{L}$ be a symmetric matrix corresponding to the square root of the Fisher information matrix, i.e. $\boldsymbol{L}'\boldsymbol{L} = \mathcal{F}(\boldsymbol{\theta})$. Then, it is possible to rewrite equations 9 and 10 as functions of $\boldsymbol{L}$, i.e.:

$$W_M = [\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0]^T \mathcal{F}(\boldsymbol{\theta}_0)[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0] = [\boldsymbol{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)]^T \underbrace{\boldsymbol{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)}_{\boldsymbol{V}_1}$$
$$S = \boldsymbol{U}^T(\boldsymbol{\theta}_0)\mathcal{F}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{U}(\boldsymbol{\theta}_0) = [(\boldsymbol{L}^{-1})^T\boldsymbol{U}(\boldsymbol{\theta}_0)]^T \underbrace{(\boldsymbol{L}^{-1})^T\boldsymbol{U}(\boldsymbol{\theta}_0)}_{\boldsymbol{V}_2}.$$

It is demonstrated that $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are two standardized vectors, which under regularity conditions have a multivariate normal asymptotic distribution, with zero expectation and variance-covariance matrices equal to the identity [19]. Then, the product of these two vectors defines the Gradient statistic [29]:

$$G = \boldsymbol{V}_2^T\boldsymbol{V}_1 = [(\boldsymbol{L}^{-1})^T\boldsymbol{U}(\boldsymbol{\theta}_0)]^T\boldsymbol{L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \mathbf{U}(\boldsymbol{\theta}_0)^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0). \tag{11}$$

This expression (11) is simpler than the Wald or Score test statistics and does not require matrix inversion, being very useful in tests associated with generalized linear models and extensions. According to Ref. [29], the gradient statistic is not necessarily non-negative and, unlike the likelihood ratio and score statistics, the G statistic is not invariant with respect to nonlinear reparameterizations of θ.

Some important properties of Gradient statistic are:

(i) If the logarithm of the likelihood function, $l(\boldsymbol{\theta})$, is unimodal and differentiable in $\boldsymbol{\theta}_0$, then: $G = \mathbf{U}(\boldsymbol{\theta}_0)^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \geq 0$, almost certainly.
(ii) If there is an unbiased estimator for $\boldsymbol{\theta}$ then $E(G) = p$, where p denotes the number of degrees of freedom relating to the restrictions under H_0.
(iii) Under conditions of regularity, it is possible to exchange the order of integration and derivation, that is: $\int \hat{\boldsymbol{\theta}} \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} d(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}^T} \int \hat{\boldsymbol{\theta}} L(\boldsymbol{\theta}) d(\boldsymbol{\theta})$.
(iv) Asymptotically, gradient statistics have a chi-squared distribution, with degrees of freedom equal to the number of restrictions given by the null hypothesis.

Other statistical properties of this test, such as power, Bartlett-type correction, robustness and applications are described by Refs [8,19,30], respectively.

3.1. Proposed gradient test

Our proposed alternative test, based on the gradient statistic, follows the same philosophy of the work developed by Ref. [15] in the following respects: (i) it is based on tests for regression coefficients and does not require estimated probabilities with hypotheses given by equation (5); and (ii) it arises from the asymptotic distribution of maximum likelihood estimators in GLMs, as defined in Section 2. However, unlike the Wald test, the gradient test does not require computing the second derivative of the log-likelihood function to evaluate the Fisher information matrix. Instead, it requires the only the first derivative. Under these conditions, considering a random sample of n individuals, k response categories, and $T-1$ transitions, we have the following likelihood function:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{b=1}^{k} \prod_{t=1}^{T-1} [\pi_{ab}(t)]^{y_{ijt}} \tag{12}$$

where y_{ijt} is an indicator function for category $b \in S$. From Equation (12) above, we can specify the first derivatives for models (1) and (2), which are presented in Appendices A and B respectively.

Then, under H_0 the gradient test statistic to assess stationarity is given by

$$G_t = \mathbf{U}(\boldsymbol{\theta}_0)^T(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}_0) \tag{13}$$

where $\mathbf{U}(\boldsymbol{\theta_0}) = \partial\ell(\boldsymbol{\theta_0})/\partial\boldsymbol{\theta_0}$ is the first derivative (score function) of the model under stationarity, and $\hat{\boldsymbol{\theta}}_t$ and $\boldsymbol{\theta}_0$ represent the estimates of the parameters under non-stationary and stationary processes, respectively. This statistic is simpler than Λ or W, as its expression does not require an inverse matrix. Asymptotically, G_t has a χ^2 distribution with degrees of freedom equivalent to the difference between the number of parameters of the non-stationary and stationary models (the same as for the LR and Wald tests).

4. Simulation study

In this section, we present a simulation study following the procedures implemented by Ref. [15]. We provide R scripts [25] to compute the gradient test and the true transition probability matrices used in the simulation studies at https://github.com/Idemauro/Gradient_test.git. The simulation process consisted of the following procedure:

(i) The data are simulated under two conditions: a) stationary (if a single stationary matrix) or non-stationary (non-homogeneous matrices, one for each time), considering a first-order Markov chain. For these we used transition matrices from Ref. [15] with $k = 5$ response categories, considering 3, 4 and 5 time occasions. We used only one matrix to simulate a stationary process and $T-1$ matrices for non-stationary processes; different matrices were used for the treatment and the control, in each case.

(ii) For each simulated dataset, models (14) and (15) were fitted assuming stationarity and non-stationarity, according to the nature of response, nominal or ordinal, respectively:

$$\eta = \log\left(\frac{\pi_{ab}(t)}{\pi_{ak}(t)}\right) = \lambda_{b(t)} + \beta_{b(t)}(\text{treatment}) + \alpha_{bt}(\text{previous answer}), \tag{14}$$

$$\eta = \log\left(\frac{\gamma_{ab}(t)}{1-\gamma_{ab}(t)}\right) = \lambda_{b(t)} + \beta_{t}(\text{treatment}) + \alpha_{t}(\text{previous answer}). \tag{15}$$

(iii) For each type of response (nominal or ordinal) 10, 000 datasets were simulated with 3 time occasions, $T = 3, 4, 5$. We also considered three different sample sizes ($N = 100, 200, 500$).

(iv) After fitting the models, the classical [3], LR, Wald [15], and gradient tests were carried out. In all scenarios, rejection rates for each test were obtained for the significance levels of 1%, 5% and 10%.

When we simulate data using a single transition probability matrix, i.e. under the true null hypothesis, we create a scenario with stationary data (Scenario 1), in which we expect the tests to keep the rejection rates within specified limits (1%, 5%, 10%), with this it is possible to study the tests' type-I error rates. On the other hand, if the data are generated using different transition probability matrices, we create non-stationarity scenario (Scenario 2), under the false null hypothesis, being it is possible, therefore, to apply the tests in order to calculate the type II error and consequently, we are able to study the tests' power functions.

The computational implementation was done using `R` software [25], using the ordinal package [5] to fit the models in both cases: ordinal and nominal. This package also was used for the nominal models because the `clm` function has the gradient statistic implemented, but it is necessary to declare the `clm` function for nominal effects. So, the transition probabilities estimated by this function are equivalent to those obtained by the function using the `multinom` function (which does not have the gradient function implemented) from the `nnet` package. Moreover, as a theoretical argument, according to Ref. [1], nominal variables can be studied as ordinal without considering the order. Furthermore, the `markovchain` package [28] was used to assist in the data simulation process.

The simulation results for the nominal case are presented in Table 1. First, under stationarity (Scenario 1), we verified that the rejection rates of the statistics were in agreement with the significance levels (10%, 5% and 1%). However, the classical test [3] was more conservative and this result is in agreement with studies carried out by Ref. [15]. Regarding the tests' power (Scenario 2), as expected, we observed that the power increases with the sample size. The proposed test (G statistic) showed the highest power for $N = 100$ and $N = 200$. For $N = 500$, all statistics are assumed to be asymptotically equivalent.

The results for ordinal case are presented in Table 2. For scenario 1, the AG, LR and Gradient statistic, returned values greater than the nominal rates for $N = 100$ and $N = 200$ and the Wald test was more conservative. When the sample size increased, the test sizes stabilized, with $N = 500$, all statistics were asymptotically equivalent. Although the proposed test has presented test sizes above those established, it was competitive. Moreover, for scenario 2, the power function of the proposed test (G) presented the highest power for $N = 100$ and $N = 200$.

Table 1. Rejection rates for nominal responses, type-I error rates (test size) and power function, for the Anderson and Goodman (AG), Wald (W), likelihood ratio (LR) and Gradient (G) tests, obtained from 10, 000 simulated datasets assuming a first-order Markov model under stationary (Scenario 1) and non-stationary (Scenario 2) processes.

		Time occasions								
		$T = 3$			$T = 4$			$T = 5$		
Levels		10%	5%	1%	10%	5%	1%	10%	5%	1%
		Scenario 1								
	AG	0.1000	0.0494	0.0080	0.0909	0.0438	0.0081	0.0845	0.0392	0.0063
$N = 100$	W	0.1079	0.0559	0.0126	0.1141	0.0622	0.0181	0.1069	0.0573	0.0158
	LR	0.1028	0.0529	0.0096	0.0956	0.0462	0.0081	0.0893	0.0408	0.0053
	G	0.1147	0.0618	0.0141	0.1181	0.0622	0.0151	0.1170	0.0615	0.0122
	AG	0.1089	0.0574	0.0117	0.1096	0.0562	0.0122	0.1114	0.0593	0.0119
$N = 200$	W	0.1125	0.0616	0.0140	0.1238	0.0686	0.0175	0.1239	0.0689	0.0193
	LR	0.1100	0.0584	0.0128	0.1138	0.0586	0.0129	0.1167	0.062	0.0133
	G	0.1163	0.0634	0.0148	0.1248	0.0675	0.0161	0.1328	0.074	0.0177
	AG	0.1045	0.0524	0.0126	0.1047	0.0555	0.0126	0.1014	0.0519	0.0104
$N = 500$	W	0.1057	0.0544	0.0133	0.1098	0.0580	0.0147	0.1045	0.0544	0.0112
	LR	0.1062	0.0533	0.0128	0.1072	0.0571	0.0134	0.1032	0.0534	0.0104
	G	0.1085	0.0549	0.0133	0.1114	0.0593	0.0139	0.1073	0.0586	0.0121
		Scenario 2								
	AG	0.6341	0.5026	0.2523	0.6814	0.5532	0.3101	0.8030	0.6840	0.4270
$N = 100$	W	0.6118	0.4694	0.2140	0.6653	0.5349	0.2881	0.7650	0.6610	0.4020
	LR	0.6530	0.5272	0.2809	0.6866	0.5644	0.3242	0.8160	0.7000	0.4540
	G	0.6769	0.5596	0.3225	0.7125	0.5978	0.3668	0.8510	0.7550	0.5290
	AG	0.9396	0.8895	0.7288	0.9701	0.942	0.8380	0.9951	0.9893	0.9502
$N = 200$	W	0.9349	0.8791	0.7041	0.9685	0.9363	0.8270	0.9942	0.9871	0.9438
	LR	0.9419	0.8952	0.7459	0.9713	0.9424	0.8402	0.9956	0.9899	0.9549
	G	0.9448	0.9012	0.7651	0.9733	0.9464	0.8509	0.9964	0.9920	0.9606
	AG	1.000	0.9998	0.9984	1.000	1.000	0.9999	1.000	1.000	1.000
$N = 500$	W	1.000	0.9997	0.9982	1.000	1.000	0.9999	1.000	1.000	1.000
	LR	1.000	0.9998	0.9985	1.000	1.000	0.9999	1.000	1.000	1.000
	G	1.000	0.9998	0.9986	1.000	1.000	0.9999	1.000	1.000	1.000

Finally, for both variables, the time occasions (T) did not influence the results and, in general, the gradient test showed a good performance, especially with respect to power.

5. Case studies

Below we present the analysis of two real datasets, one with nominal responses, and the other with ordinal responses.

5.1. Case study 1: entomology data

As an application to the nominal case, we present the analysis of a dataset originally presented in Ref. [17], that aimed to evaluate the preference of caterpillars (*Helicoverpa armigera*) for different structures of soybean and cotton plants. The experiment aimed to understand what the caterpillars attacked most in order to find better strategies for biological control in the field. The experiment was carried out at the Laboratory of Insect Biology, in the Entomology Department at the “Luiz de Queiroz” College, University of São Paulo, Brazil, 2016. The design was completely randomized with two treatments: soybean and cotton plant structures (leaves, pods, and cotton bolls). Eighty caterpillars were used, raised on an artificial diet until they reached the instar to be used in the experiments.

Table 2. Rejection rates for ordinal responses, type-I error rates (test size) and power function, for the Anderson and Goodman (AG), Wald (W), likelihood ratio (LR) and Gradient (G) tests, derived from 10,000 simulations assuming a first-order Markov model under stationary (Scenario 1) and non-stationary (Scenario 2) process.

		Time occasions								
		$T=3$			$T=4$			$T=5$		
Levels		10%	5%	1%	10%	5%	1%	10%	5%	1%
		Scenario 1								
	AG	0.1286	0.0703	0.0154	0.1345	0.0709	0.0155	0.1346	0.0734	0.0181
$N=100$	W	0.1036	0.0507	0.0107	0.0933	0.0436	0.0079	0.0841	0.0407	0.0086
	LR	0.1251	0.0679	0.0154	0.1258	0.0621	0.0113	0.1166	0.0603	0.0120
	G	0.1397	0.0791	0.0197	0.1508	0.0839	0.0204	0.1554	0.0865	0.0206
	AG	0.1151	0.0582	0.0134	0.1121	0.0569	0.0120	0.1230	0.0679	0.0153
$N=200$	W	0.1036	0.0517	0.0097	0.0902	0.0428	0.0088	0.0967	0.0486	0.0096
	LR	0.1140	0.0570	0.0128	0.1077	0.0529	0.0116	0.1169	0.0632	0.0142
	G	0.1211	0.0632	0.0147	0.1194	0.0623	0.0140	0.1366	0.0774	0.0202
	AG	0.1040	0.0545	0.0116	0.1066	0.0562	0.0130	0.1138	0.0564	0.0122
$N=500$	W	0.1000	0.0518	0.0102	0.0985	0.0516	0.0115	0.1028	0.0502	0.0089
	LR	0.1038	0.0551	0.0115	0.1048	0.0541	0.0125	0.1104	0.0552	0.0112
	G	0.1060	0.0563	0.0130	0.1104	0.0574	0.0131	0.1181	0.0601	0.0126
		Scenario 2								
	AG	0.7287	0.5958	0.3263	0.9626	0.9275	0.7918	0.9937	0.9870	0.9491
$N=100$	W	0.6457	0.4817	0.2047	0.9344	0.8737	0.6545	0.9859	0.9677	0.8756
	LR	0.7369	0.6014	0.3317	0.9615	0.9232	0.7819	0.9931	0.9849	0.9391
	G	0.7685	0.6474	0.3927	0.9681	0.9405	0.8273	0.9952	0.9891	0.9588
	AG	0.9780	0.9502	0.8300	0.9998	0.9998	0.9988	1,0000	1,0000	10,000
$N=200$	W	0.9699	0.9273	0.7609	0.9998	0.9996	0.9979	1,0000	10,000	10,000
	LR	0.9794	0.9527	0.8417	0.9998	0.9998	0.999	10,000	10,000	10,000
	G	0.9819	0.9586	0.8630	0.9998	0.9998	0.9992	10,000	10,000	10,000
	AG	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$N=500$	W	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	LR	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	G	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

For preference tests on soybeans or cotton, leaves and fruits were collected and offered simultaneously in Petri dishes. After being introduced in the Petri dishes, the caterpillars' behaviours were classified in the following categories: 'not feeding', 'feeding on the leaf', or 'feeding on the fruit', that we summarize as 1: Nothing, 2: Leaf and 3: Fruit, respectively. The behaviours were observed at four time occasions: 1 min (beginning), after 15 min, 30 min, and 45 min of exposure, characterizing a discrete-time longitudinal study with a nominal response variable. As it is an experimental study in the area of Entomology, where there are generally material restrictions, the sample size is considered satisfactory, as it is close to 100.

Figure 1 represents the distribution pattern of the response categories for the two treatments (soybean and cotton). We observe a difference in the behavioural patterns between the two treatments, with a greater preference for leaves in the soybean group and fruit in the cotton group. The behaviours also seem to change over time, hence the importance of studying the behavioural transitions. To analyse this data, we considered the model described by Equation (14), considering a first-order Markov dependence and treatment effect. We fitted models assuming stationarity and non-stationarity, and the statistics for the tests were: $\xi = 12.22(p = 0.977)$, $W = 11.27(0.999)$, $\Lambda = 12.78(p = 0.999)$ and $G = 12.61(0.999)$, all associated with 32 degrees of freedom. All tests indicate that there is no evidence supporting a non-stationary data generation process. We also found that

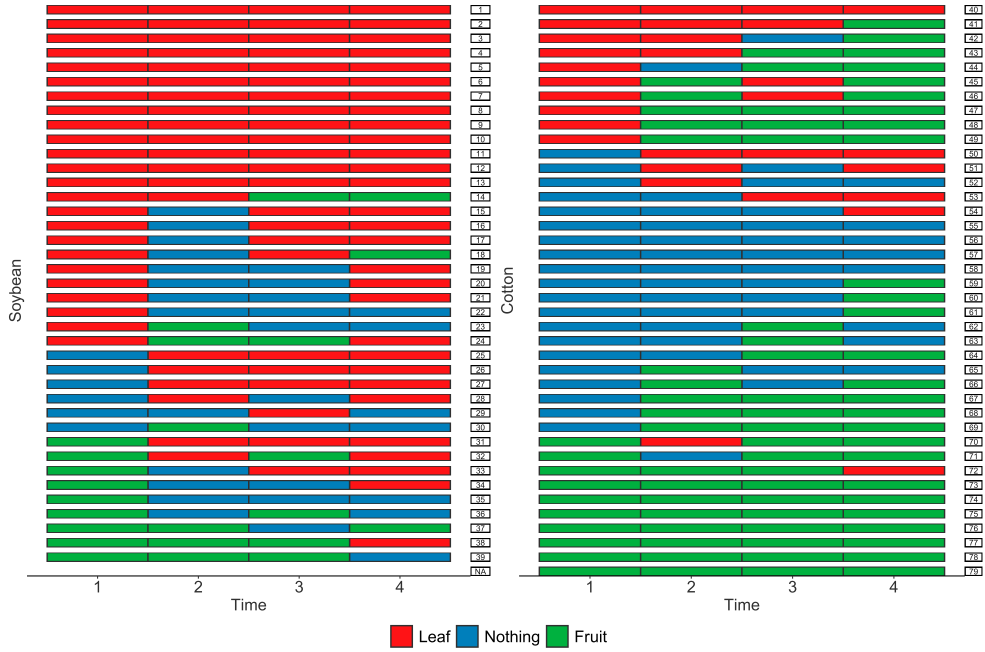

Figure 1. Observed settings of caterpillars exhibiting three behaviours: 1: Nothing, 2: Leaf and 3: Fruit, considering two treatments: soybean (left panel) and cotton (right panel).

Table 3. Parameter estimates and associated standard errors and *p*-values by Wald test for the transition model assuming stationarity, in the study of the preference of caterpillars (*Helicoverpa armigera*).

Parameters	Estimates	Standard error	*p*-value
λ_2	0.3783	0.3057	0.2159
λ_3	−0.6783	0.3793	0.0737
$\beta_2(soya)$	1.4816	0.4173	0.0004
$\beta_3(soya)$	2.4699	0.4230	<0.0001
α_{22}	−1.1886	0.5010	0.0176
α_{23}	0.5940	0.4807	0.2166
α_{32}	−2.0457	0.4388	<0.0001
α_{33}	−2.0185	0.5359	0.0002

Notes: Category 1 (Nothing) is used as the reference category, and categories 2 and 3 represent Leaf and Fruit. The λ parameters are the intercepts, the β parameters are the treatment effects, and the α parameters are the effects of the Markov covariate.

the treatment effect as well as Markov covariate were significant by likelihood ratio test ($p < 0.001$). The parameter estimates and associated standard errors for the stationary model are presented in Table 3, assuming category 1 (Nothing) as the reference category and cotton as the reference category for treatment. The results indicate that the caterpillars' preference pattern over time depends on the crop (soybean or cotton) but also on their previous choice.

The parameter estimates were used to obtain the transition probabilities (Figure 2). The transition probabilities of the three previous states for the fruit category are greater for

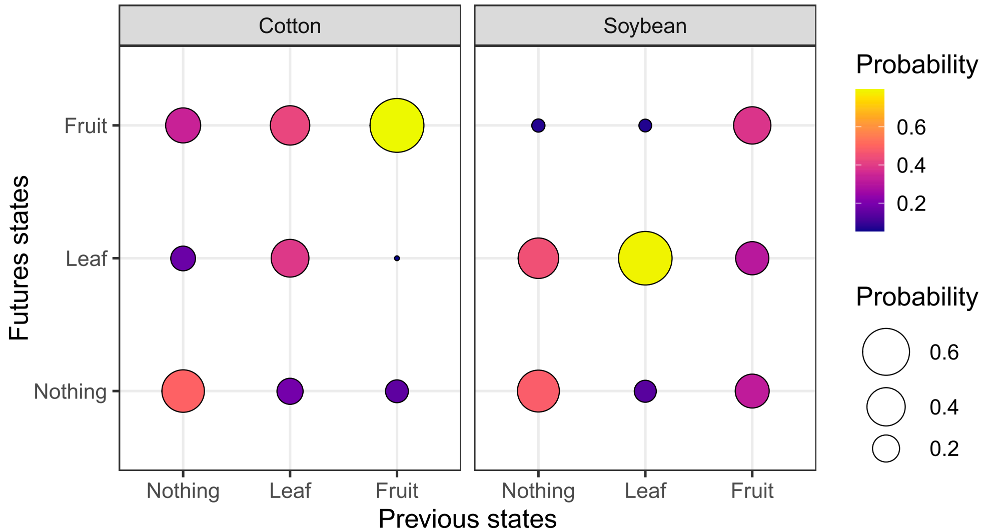

Figure 2. Estimated transition probabilities of caterpillars exhibiting three feeding behaviours: 1: Nothing, 2: Leaf and 3: Fruit, considering two treatments: soybean and cotton crops, from the stationary model.

the cotton crop. On the other hand, the probabilities of transition to the leaf category are greater for the soybean treatment. These results suggest that biological control measures should potentially be aimed at leaf structures in soybean crops and fruits in cotton crops for more successful outcomes.

5.2. Case study 2: pig behaviour data

As a second application we present a case study considered by Ref. [14] to present the central ideas of transition models and, subsequently used to apply stationarity tests (Wald and likelihood ratio) in Ref. [15]. The data refer to the degree of injury to the frontal part in pigs, that were pure genetic lineages and used for reproduction, at a commercial farm group in Brazil. In this study, an important issue is that if animals have a high degree of injury, their economic value depreciates, and consequently they are slaughtered. The experimental design was conducted to evaluate whether the use of environmental enrichment, which consisted of simple objects placed in the animal stalls, reduced aggressiveness, and, consequently, the degree of injuries. In total, 124 animals were evaluated, half of them exposed to environmental enrichment and the rest forming the control group. Injury assessments were carried out monthly from April to July/2014. They were classified in a ordinal scale: 1 – no lesions, 2 – moderate lesions and 3 – serious lesions. More details about this experiment can be found in Ref. [14].

The distribution of the degree of injury to the animals on the four study occasions is shown in Figure 3, in which greater degrees of injury were observed associated with the untreated group.

For data analysis, we followed the same procedures as in Example 1, fitting first-order transition models under stationarity and non-stationarity. Since the response is ordinal, the proportional odds model was assumed, as described by the structure given from

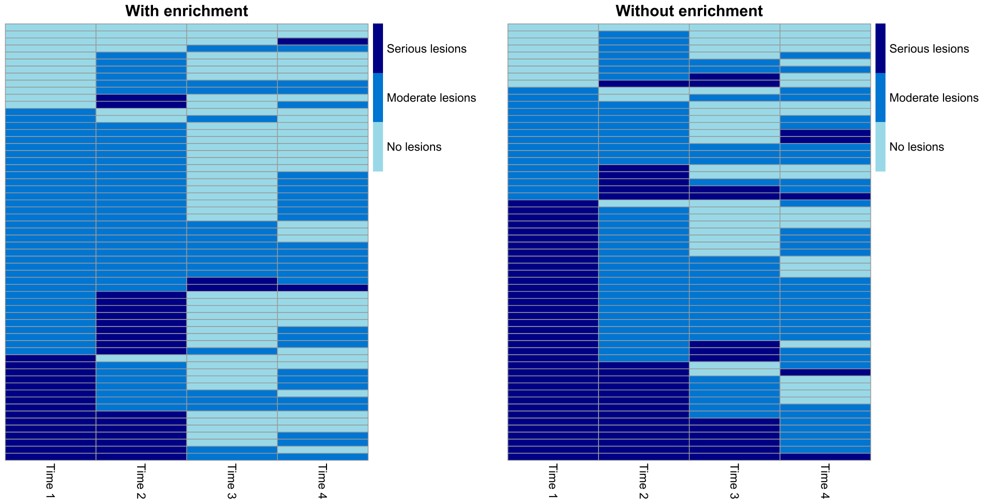

Figure 3. Observed settings of lesion degree on the front part of animals: 1 – no lesions, 2 - moderate and 3 – serious, considering the use or not of environmental enrichment, in four time occasions.

Equation (2), in which $\boldsymbol{\beta}$ is associated with the environmental enrichment effect and $\boldsymbol{\alpha}$ with the Markov covariate (previous response). Unlike the first study, here the tests were in agreement with the non-stationarity hypothesis ($p < 0.01$). The statistics for the tests were: $\xi = 78.32$, $W = 69.04$, $\Lambda = 82.32$ and $G = 88.98$, associated with 10 degrees of freedom. Therefore, we assume that the process is non-stationary and can be represented by three models, one for each time transition, whose estimated values for the parameters are presented in Table 4.

It was observed that the effect of environmental enrichment was not significant in the first and third transitions, justified by the fact that initially the animals were still learning to use the objects and, later, may have lost interest. It is also observed that in these transitions the degree of injury was influenced by the Markov covariate. Using the coefficients of these models, the transition probabilities can be estimated, which are presented in Figure 4. It can be observed that the highest probabilities of transition to state 1 (without injury) occur in the treated group, especially in the second transition.

6. Final considerations

Transition models are an alternative for the analysis of longitudinal categorical data, especially when the main interest is not in the description of the serial correlation as in the classical marginal and mixed models, but in the prediction and verification of the contribution of the covariates of the experimental design.

In this context, it is necessary to verify the stationarity condition in terms of homogeneity of the transition probabilities. As this condition is not always true, it is critical to employ tests to assess it. We presented an alternative test based on the gradient statistic, whose performance under simulation was satisfactory regarding power and type-I error rates, and competitive with other tests previously described in the literature. In addition to its competitors, this test is easily applicable since its statistic involves only the first

Table 4. Parameter estimates and associated standard errors and p-values by Wald test for the three transitions in the study with degree of injury in pigs, considering 1 (no lesion) and E1: with environmental enrichment as the reference categories; E2: without environmental enrichment.

Parameters	Estimate	Standard error	p-value
	First transition		
$\lambda_{2(1)}$	−1.3016	0.5162	0.0116
$\lambda_{3(1)}$	2.0675	0.5564	0.0002
$\beta_{(1)}(E2)$	0.0249	0.3950	0.9496
$\alpha_{2(1)}$	1.1032	0.5723	0.0539
$\alpha_{3(1)}$	1.6827	0.5851	0.0040
	Second transition		
$\lambda_{2(2)}$	1.6121	0.7184	0.0248
$\lambda_{3(2)}$	3.6580	0.7825	0.0000
$\beta_{(2)}(E2)$	1.3084	0.3708	0.0004
$\alpha_{2(2)}$	0.7799	0.7317	0.2865
$\alpha_{3(2)}$	1.1929	0.7682	0.1204
	Third transition		
$\lambda_{2(3)}$	0.2574	0.2824	0.3622
$\lambda_{3(3)}$	3.5418	0.4932	0.0000
$\beta_{(3)}(E2)$	0.3757	0.3814	0.3247
$\alpha_{2(3)}$	0.5053	0.3943	0.1999
$\alpha_{3(3)}$	1.5696	0.6655	0.0183

Note:The λ parameters are the intercepts, the β parameters are the treatment effects, and the α parameters are the effects of the Markov covariate.

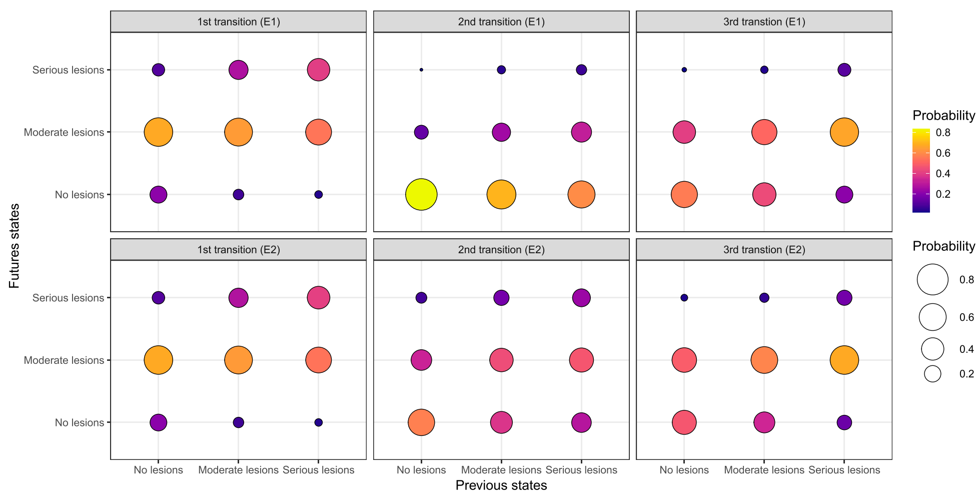

Figure 4. Estimated transition probabilities of lesion degree on the front part of animals, according to non-stationary first-order transition models: 1 – no lesions, 2 – moderate lesions and 3 – serious lesions, considering the use (E1) or not (E2) of environmental enrichment.

derivative of the log-likelihood. However, there is a need for additional studies focussed on more complex experimental designs, such as factorial treatment schemes. With a greater number of covariates, transition models tend to be overparameterized and the larger sample size avoids sparse data, which compromises the parametric estimation of any process.

Acknowledgments

This work has partially emanated from LVTP's doctoral thesis "Transition models with categorical data, in processes of discrete and continuous time, with applications to entomology".

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was funded by the Brazilian Foundation "Coordenação de Aperfeiçoamento de Pessoal de Nível Superior" (CAPES) process numbers 88882.378347/2019-01 and 88887.716582/2022-00.

ORCID

Rafael de Andrade Moral http://orcid.org/0000-0002-0875-3563

References

[1] A. Agresti, *Categorical Data Analysis*, 3rd ed., Wiley, New York, 2012.
[2] P.S. Albert, *A transitional model for longitudinal binary data subject to nonignorable missing data*, Biometrics 56 (2000), pp. 602–608.
[3] T.W. Anderson and L.A. Goodman, *Statistical inference about Markov chains*, Ann. Math. Stat. 28 (1957), pp. 89–110.
[4] F. Bartolucci, A. Farcomeni, and F. Pennoni, *Latent Markov Models for Longitudinal Data*, CRC Press, New York, 2013.
[5] R.H.B. Christensen, *Cumulative link models for ordinal regression with the R package ordinal*, J. Stat. Softw. 35 (2018), pp. 1–46.
[6] P.J. Diggle, P. Heagerty, P.J. Heagerty, K.Y. Liang, and S. Zeger, *Analysis of Longitudinal Data*, Oxford University Press, New York, 2002.
[7] L.D. Fernandes, A.S. Mata, W.A.C. Godoy, and C. Reigada, *Refuge distributions and landscape connectivity affect host-parasitoid dynamics: Motivations for biological control in agroecosystems*, PLoS One 17 (2022), pp. e0267037.
[8] S. Ferrari, E.C. Pinheiro, and S. Zeger, *Small-sample likelihood inference in extreme-value regression models*, J. Stat. Comput. Simul. 84 (2014), pp. 582–595.
[9] M. Ganjali, Z. Rezaee, and S. & Zeger, *A transition model for analysis of repeated measure ordinal response data to identify the effects of different treatments*, Drug Inf. J. 41 (2007), pp. 527–534.
[10] T. Hayakawa and M.L. Puri, *Asymptotic expansions of the distribuition os some test statistics*, Ann. Inst. Stat. Math. 37 (1985), pp. 95–108.
[11] P.J. Heagerty, *Marginalized transition models and likelihood inference for longitudinal categorical data*, Biometrics 58 (2002), pp. 342–351.
[12] D.W. Hosmer, Jr., S. Lemeshow, and R.X. Sturdivant, *Applied Logistic Regression*, John Wiley & Sons, New Jersey, 2013.
[13] S. Karlin and H.M. Taylor, *A First Course in Stochastic Processes*, Academic Press, New York, 1975.
[14] I.A.R. de Lara, J.P. Hinde, A.C. de Castro, and I.J.O. da Silva, *A proportional odds transition model for ordinal responses with an application to pig behaviour*, J. Appl. Stat. 44 (2017), pp. 1031–1046.
[15] I.A.R. de Lara, J.P. Hinde, and C.A. Taconeli, *An alternative method for evaluating stationarity in transition models*, J. Stat. Comput. Simul. 87 (2018), pp. 2962–2980.
[16] I.A.R. de Lara, R.A. Moral, C.A. Taconeli, C. Reigada, and L. Hinde, *A generalized transition model for grouped longitudinal categorical data*, Biom. J. 62 (2020), pp. 1837–1858.

[17] I.A.R. de Lara, C. Reigada, and C.A. Taconeli, *Transition models applied to interactions involving agricultural pests*, in *Modelling Insect Populations in Agricultural Landscapes. Entomology in Focus*, Vol. 8, R.A. Moral and W.A. Godoy, eds., Springer, Cham, 2023. Available at: https://doi.org/10.1007/978-3-031-43098-5-5.
[18] K. Lee and M.J. Daniels, *A class of Markov models for longitudinal ordinal data*, Biometrics 63 (2007), pp. 1060–1067.
[19] A.J. Lemonte, *The Gradient Test: Another Likelihood-Based Test*, Academic Press, London, 2016.
[20] A.J. Lemonte and S.L.P. Ferrari, *The local power of the gradient test*, Ann. Inst. Stat. Math. 64 (2012), pp. 373–381.
[21] P. McCullagh, *Regression models for ordinal data*, J. R. Stat. Soc. Ser. B: Stat. Methodol. 42 (1980), pp. 109–127.
[22] G. Molenberghs and G. Verbeke, *Models for Discrete Longitudinal Data*, Springer Science & Business Media, New York, 2006.
[23] J.S. Oliveira, C. Reigada, A.J.F. Diniz, I.A.R. de Lara, R.A. Moral, and J.R.P. Parra, *Can parasitism by Tamarixia radiata (Hymenoptera: Eulophidae) affect the movement and oviposition behavior of Diaphorina citri (Hemiptera: Psyllidae)?* J. Insect Behav. 35 (2022), pp. 183–194; Available at: https://doi.org/10.1007/s10905-022-09811-6.
[24] C.R. Rao, *Score tests: Historical review and recent developments*, in *Advances in Ranking and Selection, Multiple Comparisons and Reliability*, N. Balakrshnan, N. Kannan, and H.N. Nagaraja, eds., Birkhuser, Boston, MA, 2005.
[25] R Core Team, *A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2022. Available at: https://www.R-project.org/.
[26] I.A.R. de Lara, J. Hinde, and C.A. Taconeli, *Global and local tests to assess stationarity of Markov transition models*, Commun. Stat. Simul. Comput. 48 (2018), pp. 1–21; Available at: https://doi.org/10.1080/03610918.2017.1406504.
[27] S.M. Ross, *Stochastic Processes*, Vol. 2, John Wiley & Sons, New York, 1996.
[28] G.A. Spedicato, T.S. Kang, S.B. Yalamanchi, D. Yadav, and I. Cordón, *The markovchain package: A package for easily handling Discrete Markov Chains in R*, Vienna, Austria. Available at: https://www.R-project.org/.
[29] G.R. Terrell, *The gradient statistic*, Comput. Sci. Stat. 34 (2002), pp. 206–215.
[30] T.M. Vargas, S.L. Ferrari, and A.J. Lemonte, *Improved likelihood inference in generalized linear models*, Comput. Stat. Data Anal. 74 (2014), pp. 110–124.
[31] J.H. Ware, S. Lipsitz, and F.E. Speizer, *Issues in the analysus of repeated categorical outcomes*, Stat. Med. 7 (1988), pp. 95–107.
[32] F. Yu, H. Morgenstern, E. Hurwitz, and T.R. Berlin, *Use of a Markov transition model to analyse longitudinal low-back pain data*, Stat. Meth. Med. Res. 12 (2003), pp. 321–331.
[33] L. Zeng and R.J. Cook, *Transition models for multivariate longitudinal binary data*, J. Am. Stat. Assoc. 102 (2007), pp. 211–223.

Appendices

Here we present calculations of the derivatives of the log-likelihood functions for both the nominal and ordinal cases.

Appendix 1. Nominal case

Logarithm of the likelihood function:

$$\begin{aligned}\ell(\boldsymbol{\theta}) &= \sum_{i=1}^{n}\sum_{b=1}^{k}\sum_{t=1}^{T-1}\log\{[\pi_{ab}(t)]^{y_{ijt}}\} = \sum_{i=1}^{n}\sum_{b=1}^{k}\sum_{t=1}^{T-1}\{y_{ijt}\log[\pi_{ab}(t)]\}\\ &= \sum_{i=1}^{n}\sum_{t=1}^{T-1}\left\{\sum_{b=1}^{k-1} y_{ijt}\log\left[\frac{\exp(\lambda_{ab(t)}+\boldsymbol{\delta}_t^T\mathbf{x}^*)}{1+\sum_{b=1}^{k-1}\exp(\lambda_{ab(t)}+\boldsymbol{\delta}_t^T\mathbf{x}^*)}\right]\right\}\end{aligned}$$

$$= \sum_{i=1}^{n}\sum_{t=1}^{T-1}\left\{\sum_{b=1}^{k-1} y_{ijt}\log\{\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)\} - \log\left\{1 + \sum_{b=1}^{k-1}\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)\right\}\right\}$$

$$= \sum_{i=1}^{n}\sum_{t=1}^{T-1}\left\{\sum_{b=1}^{k-1} y_{ijt}(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*) - \log\left\{1 + \sum_{b=1}^{k-1}\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)\right\}\right\}.$$

First derivative:

$$\mathbf{U}(\lambda_b) = \sum_{i=1}^{n}\sum_{t=1}^{T-1}\left\{\sum_{b=1}^{k-1} y_{ijt} - \frac{\sum_{b=1}^{k-1}\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)}{1 + \sum_{b=1}^{k-1}\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)}\right\}$$

$$\mathbf{U}(\boldsymbol{\delta}_t) = \sum_{i=1}^{n}\sum_{t=1}^{T-1}\left\{\sum_{b=1}^{k-1} y_{ijt}\mathbf{x}^* - \frac{\sum_{b=1}^{k-1}\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)\mathbf{x}^*}{1 + \sum_{b=1}^{k-1}\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)}\right\}.$$

Appendix 2. Ordinal case

Logarithm of the likelihood function:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{b=1}^{k}\sum_{t=1}^{T-1}\log\{[\pi_{ab}(t)]^{y_{ijt}}\} = \sum_{i=1}^{n}\sum_{b=1}^{k}\sum_{t=1}^{T-1}\{y_{ijt}\log[\pi_{ab}(t)]\}$$

$$= \sum_{i=1}^{n}\sum_{b=1}^{k}\sum_{t=1}^{T-1}\{y_{ijt}\log[\gamma_{ab}(t) - \gamma_{a(b-1)}(t)]\}$$

$$= \sum_{i=1}^{n}\sum_{b=1}^{k}\sum_{t=1}^{T-1}\left\{y_{ijt}\log\left[\frac{\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)}{1 + \exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)} - \frac{\exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)}{1 + \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)}\right]\right\}$$

$$= \sum_{i=1}^{n}\sum_{b=1}^{k}\sum_{t=1}^{T-1}\left\{y_{ijt}\log\left[\frac{A - B}{[1 + \exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)][1 + \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)]}\right]\right\}$$

where

$$\begin{aligned} A &= \exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)[1 + \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)] \\ &= \exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*) + \exp(\lambda_{b(t)} + \lambda_{(b-1)(t)} + 2\boldsymbol{\delta}_t^T\mathbf{x}^*) \\ B &= \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)[1 + \exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)] \\ &= \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*) + \exp(\lambda_{b(t)} + \lambda_{(b-1)(t)} + 2\boldsymbol{\delta}_t^T\mathbf{x}^*) \end{aligned}$$

and therefore

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\sum_{b=1}^{k}\sum_{t=1}^{T-1}\left\{y_{ijt}\log\left[\frac{\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*) - \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)}{[1 + \exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)][1 + \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)]}\right]\right\}$$

$$\begin{aligned} &= \sum_{i=1}^{n}\sum_{b=1}^{k}\sum_{t=1}^{T-1} y_{ijt}\{\log[\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*) - \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)] \\ &\quad - \log[1 + \exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)] - \log[1 + \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)]\}. \end{aligned}$$

First derivative:

$$\mathbf{U}(\lambda_b) = \sum_{i=1}^{n}\sum_{b=1}^{k}\sum_{t=1}^{T-1} y_{ijt}\left\{\frac{\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)}{\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*) - \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T\mathbf{x}^*)}\right.$$

$$- \frac{\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)}{1 + \exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)} \Bigg\}$$

$$\mathbf{U}(\boldsymbol{\delta}_t) = \sum_{i=1}^{n} \sum_{b=1}^{k} \sum_{t=1}^{T-1} y_{ijt} \left\{ \frac{\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)\mathbf{x}^* - \exp(\lambda_{(b-1)(t)} + \delta_t^T \mathbf{x}^*)\mathbf{x}^*}{\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*) - \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)} - \right.$$

$$\left. - \frac{\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)\mathbf{x}}{1 + \exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)} - \frac{\exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)\mathbf{x}}{1 + \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)} \right\}$$

$$= \sum_{i=1}^{n} \sum_{b=1}^{k} \sum_{t=1}^{T-1} y_{ijt} \mathbf{x}^* \left\{ 1 - \frac{\exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)}{1 + \exp(\lambda_{b(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)} - \frac{\exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)}{1 + \exp(\lambda_{(b-1)(t)} + \boldsymbol{\delta}_t^T \mathbf{x}^*)} \right\}.$$