



Exact significance test for Markov order



S.D. Pethel^{a,*}, D.W. Hahs^b

^a U.S. Army RDECOM, RDMR-WDS-W, Redstone Arsenal, AL 35898, USA

^b Torch Technologies, Inc., Huntsville, AL 35802, USA

HIGHLIGHTS

- We give a hypothesis test of Markov order from time series data.
- A novel and efficient algorithm for generating surrogate data is introduced.
- Surrogate data enables exact computation of null hypothesis p -values even for small data sets.

ARTICLE INFO

Article history:

Received 1 March 2013

Received in revised form

17 September 2013

Accepted 22 November 2013

Available online 1 December 2013

Communicated by A.C. Newell

Keywords:

Markov order

Hypothesis test

Whittle's formula

ABSTRACT

We describe an exact significance test of the null hypothesis that a Markov chain is n th order. The procedure utilizes surrogate data to yield an exact test statistic distribution valid for any sample size. Surrogate data are generated using a novel algorithm that guarantees, per shot, a uniform sampling from the set of sequences that exactly match the n th order properties of the observed data. Using the test, the Markov order of Tel Aviv rainfall data is examined.

Published by Elsevier B.V.

1. Introduction

It often happens that it is useful to describe a process as a set of discrete states with probabilistic transitions. Examples abound in various fields such as the study of chemical processes [1], DNA sequences [2], finance [3], and nonlinear dynamics [4], among others. If the transition probability to the next state is conditioned only on the present state we call this model a Markov chain. An n th-order Markov chain is a generalization that includes the past n states in the transition probability. When the conditional probabilities are not otherwise given, they are estimated from a time series of observations.

If the order of the Markov chain is in question, there are various tests and criteria available to narrow down the options. The classical approach is to formulate the question as a hypothesis test that a chain is n th order and specify a discriminating test statistic [5]. Often the chosen test statistic has a known limiting distribution, such as χ^2 , from which a p -value is calculated [6]. The use of a limiting distribution is inexact because it is only attained for infinite

data sets. For small data sets it may be a very poor approximation. In this article we develop a procedure to compute p -values that are exact, even for small data sets.

The p -value is the probability, assuming the null hypothesis, of the test statistic attaining its observed value or one more extreme. It is not the probability of the null hypothesis being correct. While a very small p -value leads one to reject the null hypothesis, a large p -value only implies that the data is consistent with the null hypothesis, not that the null hypothesis should be accepted. In addition, the significance threshold for rejection is entirely up to the user to decide.

There are alternative approaches that provide a positive answer to order selection. The Akaike information criterion (AIC), the Bayesian information criterion (BIC), and others [7–10] produce rankings over multiple orders based on their relative likelihoods and have built-in penalties for over-fitting. Unfortunately, these approaches rely on approximations that are only valid in the limit of large samples. In small sample situations one cannot be sure of their efficacy.

In spite of the shortcomings of hypothesis tests, there is an advantage in that it is possible, in principle, to perform an exact Markov order significance test that is valid for any sample size. Instead of relying on asymptotic properties, the test statistic distribution is discovered by sampling from the set of sequences (referred

* Corresponding author. Tel.: +1 2568429734.

E-mail addresses: shawn.pethel@us.army.mil, shawn.d.pethel.civ@mail.mil (S.D. Pethel), daniel.hahs@us.army.mil (D.W. Hahs).

to here as *surrogates*) that *exactly* match the n th order properties of the observed time series [11]. The challenge is in efficiently generating a large number of such samples, especially for higher orders. To our knowledge no solution to this problem has been reported in the literature.

The contribution of this work is a surrogate data procedure that has ideal properties: each sample surrogate has exactly the same transition counts as the observed sequence, one sample is generated per shot, samples are uniformly selected from the set of all possible surrogate sequences, computation time increases linearly with the length of the sequence, and any order can be accommodated. Armed with this new procedure it is now straightforward to compute the p -value of a Markov order null hypothesis exactly for any size data set. The obtained value may be used for significance testing in the standard way (as we do here) or as part of a more general procedure for model selection.

We first describe how to do hypothesis testing of Markov order using the χ^2 statistic, for which the distribution is known in the large sample limit. Next we describe the method of surrogate data generation based on Whittle's formula. Then we compare the χ^2 statistic in large and small sample cases using both the asymptotic distribution and the exact distribution obtained from the surrogates. Finally, we analyze Tel Aviv rainfall data in which AIC and BIC give differing results and discuss the use of p -values in this situation.

2. Testing Markov order using the χ^2 statistic

A sequence of observations $\mathbf{x} = \{x_1 \dots x_N\}$ form a Markov chain of order n if the conditional probability satisfies

$$p(x_{t+1}|x_t, x_{t-1} \dots) = p(x_{t+1}|x_t \dots x_{t-n+1}), \quad (1)$$

for all $n < t \leq N$. For convenience we will label the states each measurement can take by positive integers up to S . A sequence of discrete measurements may come from a process that is naturally discrete, such as a DNA sequence, or from a continuous process that has been discretized by an analog-to-digital measuring device. Unless otherwise specified a Markov process is assumed to be first order ($n = 1$). This means that the transition probabilities to a future state depend only on the present state and not on prior states. An n th order process can always be cast as first order by grouping the present state with the relevant past states into a *word*, in which case the number of states can be up to S^n . A process that has no dependence on past or present (such as a random iid process) is said to be zeroth order.

To perform an n th order null hypothesis test it is necessary to compute the distribution of a suitable higher order statistic. If the observed higher order statistic is sufficiently unlikely, then the null hypothesis is rejected. The probability, given the null hypothesis, of the test statistic attaining the observed value or one more extreme is referred to as the p -value. Typically, a p -value less than or equal to 0.05 is taken as grounds to reject the null hypothesis.

Let us begin with the assumption that $\{x_t\}$ is an observed sequence from a first order Markov process ($n = 1$) and calculate the p -value of a second order statistic using a χ^2 distribution. The null hypothesis is

$$p(x_{t+1} = i|x_t = j, x_{t-1} = k) = p(x_{t+1} = i|x_t = j), \quad (2)$$

or using Bayes' rule

$$p(x_{t+1} = i, x_t = j, x_{t-1} = k) = \frac{p(x_{t+1} = i, x_t = j)p(x_t = j, x_{t-1} = k)}{p(x_t = j)}. \quad (3)$$

The l.h.s. of (3) multiplied by $N - 2$ is the expected number of times the word $(x_{t+1} = i, x_t = j, x_{t-1} = k)$ appears in the data given the null hypothesis. The quantities on the r.h.s. are not expected values; they are taken from the observed sequence. Let E_w be the expected word count where $\sum E_w = N - 2$ and w indexes the set of all words for which the expected count is greater than zero. Similarly, let $O_w \geq 0$ be the corresponding count from the observed data. We can now define the observed χ^2 test statistic

$$\chi_{\text{obs}}^2 = \sum_w \frac{(E_w - O_w)^2}{E_w}, \quad (4)$$

which is a measure of the deviation of the observed count from the expected. The first order assumption does not uniquely determine the second order statistics; there is some freedom for χ^2 to vary from shot to shot even assuming the null hypothesis. The advantage of the χ^2 statistic is that, given the degrees of freedom d , the distribution $f(\chi^2; d)$ is known in the limit $N \rightarrow \infty$. The p -value is then obtained by integrating $f(\chi^2)$ over $\chi^2 \geq \chi_{\text{obs}}^2$.

An issue requiring some discussion is how to compute the degrees of freedom d needed to determine the χ^2 distribution. To test the n th-order hypothesis we count the observed $k = n + 1$ length words and compute the expected $k + 1$ length words. Assuming all S^k length k words are present in the data, let F be a $S^k \times S^k$ transition count matrix. The (i, j) th entry of F is the number of times word i transitions to word j . Because consecutive words overlap and differ by only one symbol, there are at most S nonzero entries in each row and column of F . It is helpful to rearrange F into block diagonal form with $kS \times S$ blocks. In each block both the rows and columns must add up to the corresponding length k word count. Taking into account row and column dependencies leaves us with $S^{k-1}(S - 1)^2$ degrees of freedom for $k > 0$, and $(S - 1)^2$ for $k = 0$ [12]. In the case that not all length k words are present in the observed data, F will be smaller than $S^k \times S^k$ and the blocks along the diagonal may be of differing size. If the size of the b th block is $r_b \times c_b$, then the total number of degrees of freedom d is $\sum (r_b - 1)(c_b - 1)$.

3. Exact test using surrogate data

The hypothesis test as described above is not exact; it relies on the χ^2 distribution valid in the asymptotic limit of infinite data. To discover the exact distribution for finite data it is necessary to evaluate χ^2 for all possible sequences that satisfy the null hypothesis. For the first order hypothesis these sequences have exactly the same joint probabilities shown on the r.h.s. of (3). Let F_{ij} be the number of times word i transitions to word j in \mathbf{x} , and let $\Gamma(\mathbf{x})$ represent the set of sequences with the same F and the same beginning and end states as the observed sequence \mathbf{x} .

The number of sequences that have the word transition count F and begin with state u and end with state v is given by Whittle's formula [13]:

$$N_{uv}(F) = \frac{\prod_i F_i!}{\prod_{ij} F_{ij}!} C_{vu} \quad (5)$$

where F_i is the sum of row i and C_{vu} is the (v, u) th cofactor of the matrix

$$F_{ij}^* = \begin{cases} \delta_{ij} - F_{ij}/F_i & \text{if } F_i > 0, \\ \delta_{ij} & \text{if } F_i = 0. \end{cases} \quad (6)$$

As an example, consider the following sequence of twelve binary observations:

$$\mathbf{x} = \{0111010111001\}. \quad (7)$$

Download English Version:

<https://daneshyari.com/en/article/8256498>

Download Persian Version:

<https://daneshyari.com/article/8256498>

[Daneshyari.com](https://daneshyari.com)