IMPACT OF COVID-19 PANDEMIC IN HEALTH SERVICES USAGE

Amanda Fernández-Fontelo, Pedro Puig, Montserrat Guillén and David Moriña

UNIVERSITAT DE BARCELONA
Facultat d'Economia i Empresa

Fundación MAPFRE

17 June 2022

Motivation

- There is an enormous global concern around 2019-novel coronavirus (SARS-CoV-2) infection in the last months, leading the World Health Organization (WHO) to declare public health emergency in early 2020
- The consequences derived from the pandemic caused by this virus have had a profound effect on many areas of human activity
- In addition to the direct consequences, in 2020 a decrease in use of health services has been detected, both those belonging to the Public Health System and services associated with private health insurances

Motivation

- The question is to know if, either due to the effect of postponing visits or due to the consequences of having suffered the virus (persistent Covid or secondary effects), there will be an excess of claims in 2022 and the following years
- There is already evidence of a higher frequency of use of Health services in the Public System but it is difficult to determine if the highest frequency of claims that will be observed will be equal to or greater than the infra-loss rate that was observed during the pandemic period

INAR models

Let X_n be a process defined by

$$X_n = \alpha_1 \circ X_{n-1} + ... + \alpha_k \circ X_{n-k} + W_n, \qquad (1)$$

where $0 < \alpha_1 < ... < \alpha_k$ and W_n is assumed to follow a Poisson distribution with a fixed mean λ.
X_n and W_n are assumed to be independent at any time n.

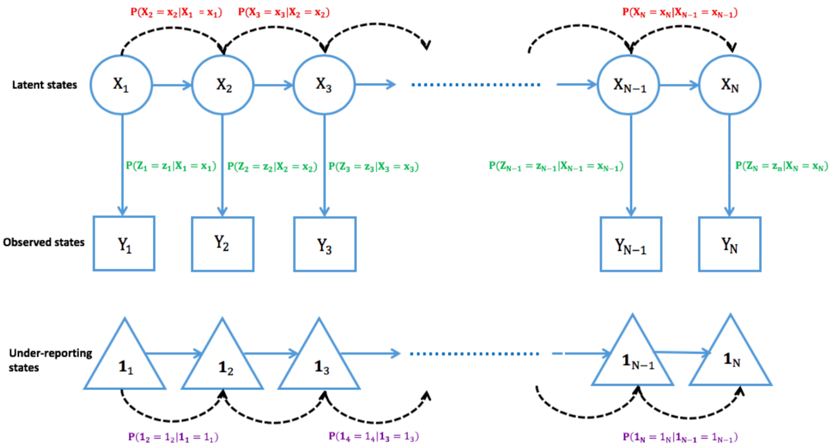
INAR models

The $\circ$ operator, called *binomial thinning*, is defined as follows:

$$p_j \circ X_{n-j} = \sum_{i=1}^{X_{n-j}} Y_i, \tag{2}$$

where Y_i are independent and identically distributed Bernoulli random variables with a probability of success equal to p_j. Therefore, if $X_{n-j} = x_{n-j}$, then $p_j \circ X_{n-j}$ is binomially distributed, with number of successes equal to x_{n-j}.

Visual model



Previously proposed models

Independent under-reporting states

Research Article

Statistics in Medicine

Received 17 February 2016, Accepted 9 June 2016 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.7026

Under-reported data analysis with INAR-hidden Markov chains

Amanda Fernández-Fontelo,[a*†] Alejandra Cabaña,[a] Pedro Puig[a] and David Moriña[b,c]

In this work, we deal with correlated under-reported data through INAR(1)-hidden Markov chain models. These models are very flexible and can be identified through its autocorrelation function, which has a very simple form. A naïve method of parameter estimation is proposed, jointly with the maximum likelihood method based on a revised version of the forward algorithm. The most-probable unobserved time series is reconstructed by means of the Viterbi algorithm. Several examples of application in the field of public health are discussed illustrating the utility of the models.

Keywords: discrete time series; emission probabilities; integer-autoregressive models; thinning operator; under-recorded data

Previously proposed models

Serially dependent under-reporting states

RESEARCH ARTICLE

WILEY Statistics in Medicine

Untangling serially dependent underreported count data for gender-based violence

Amanda Fernández-Fontelo[1,2] | **Alejandra Cabaña**[2] | **Harry Joe**[3] | **Pedro Puig**[1,4] | **David Moriña**[4]

[1]School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany

[2]Departament de Matemàtiques, Universitat Autònoma de Barcelona, Barcelona, Spain

[3]Department of Statistics, University of British Columbia, Vancouver, Canada

[4]Barcelona Graduate School of Mathematics, Departament de Matemàtiques, Universitat Autònoma de Barcelona, Bellaterra, Spain

Correspondence
Amanda Fernández-Fontelo, School of Business and Economics, Humboldt-Universität zu Berlin, 10178 Berlin, Germany; or Departament de Matemàtiques, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain.
Email: fernanda@hu-berlin.de

Underreporting in gender-based violence data is a worldwide problem leading to the underestimation of the magnitude of this social and public health concern. This problem deteriorates the data quality, providing poor and biased results that lead society to misunderstand the actual scope of this domestic violence issue. The present work proposes time series models for underreported counts based on a latent integer autoregressive of order 1 time series with Poisson distributed innovations and a latent underreporting binary state, that is, a first-order Markov chain. Relevant theoretical properties of the models are derived, and the moment-based and maximum-based methods are presented for parameter estimation. The new time series models are applied to the quarterly complaints of domestic violence against women recorded in some judicial districts of Galicia (Spain) between 2007 and 2017. The models allow quantifying the degree of underreporting. A comprehensive discussion is presented, studying how the frequency and intensity of underreporting in this public health concern are related to some interesting socioeconomic and health indicators of the provinces of Galicia (Spain).

Previously proposed models

Non-stationary processes

PLOS ONE

RESEARCH ARTICLE

Estimating the real burden of disease under a pandemic situation: The SARS-CoV2 case

Amanda Fernández-Fontelo[1]*, David Moriña[2,3,4], Alejandra Cabaña[2], Argimiro Arratia[5], Pere Puig[2]

1 Chair of Statistics, School of Business and Economics, Humboldt-Universität zu Berlin, Berlin, Germany, 2 Departament de Matemàtiques, Barcelona Graduate School of Mathematics (BGSMath), Universitat Autònoma de Barcelona, Barcelona, Spain, 3 Department of Econometrics, Statistics and Applied Economics, Riskcenter-IREA, Universitat de Barcelona, Barcelona, Spain, 4 Centre de Recerca Matemàtica (CRM), Barcelona, Spain, 5 Department of Computer Science, Universitat Politècnica de Catalunya, Barcelona, Spain

* fernanda@hu-berlin.de

Abstract

The present paper introduces a new model used to study and analyse the severe acute respiratory syndrome coronavirus 2 (SARS-CoV2) epidemic-reported-data from Spain. This is a Hidden Markov Model whose hidden layer is a regeneration process with Poisson immigration, Po-INAR(1), together with a mechanism that allows the estimation of the under-reporting in non-stationary count time series. A novelty of the model is that the expectation of the unobserved process's innovations is a time-dependent function defined in such a way that information about the spread of an epidemic, as modelled through a Susceptible-Infectious-Removed dynamical system, is incorporated into the model. In addition, the parameter controlling the intensity of the under-reporting is also made to vary with time to adjust to possible seasonality or trend in the data. Maximum likelihood methods are used to estimate the parameters of the model.

OPEN ACCESS

Citation: Fernández-Fontelo A, Moriña D, Cabaña A, Arratia A, Puig P (2020) Estimating the real burden of disease under a pandemic situation: The SARS-CoV2 case. PLoS ONE 15(12): e0242956. https://doi.org/10.1371/journal.pone.0242956

Editor: Paul K. Newton, University of Southern California, UNITED STATES

Received: July 5, 2020

Accepted: November 12, 2020

The fattening-thinning operator

Let X_n be a latent process following an INAR(1) structure such that: $X_n = \alpha \circ X_{n-1} + Z_n$, where $\mathrm{E}(X_n) = \mu_X$ and $\mathrm{Var}(X_n) = \sigma_X^2$ are the expectation and variance of X_n, respectively. Assume, for now, that $Z_n \sim \mathrm{Poisson}(\lambda)$. Let Y_n be an observed and potentially over- or under-reporting process such that:

$$Y_n = \begin{cases} X_n & 1-\omega \\ \theta \lozenge X_n & \omega, \end{cases} \tag{3}$$

The fattening-thinning operator

$\Diamond$ is the fattening-thinning operator in the sense that:

$$\theta \Diamond X_n | X_n = x_n = \sum_{j=1}^{x_n} W_j, \tag{4}$$

where W_j are i.i.d random variables defined by the following probability mass function (pmf):

$$\mathbb{P}(W_j = k|\phi_1, \phi_2) = \begin{cases} 1 - \phi_1 - \phi_2 & \text{if } k = 0 \\ \phi_1 & \text{if } k = 1 \\ \phi_2 & \text{if } k = 2 \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $\theta = (\phi_1, \phi_2)$

Misreported data

- To distinguish between under-reporting, no misreporting or over- reporting, the following can easily be computed once the model is estimated:

Under-reporting vs over-reporting

$$\begin{aligned} \text{under-reporting:} \quad & \phi_1 + 2\phi_2 < 1 \\ \text{no misreporting:} \quad & \phi_1 + 2\phi_2 = 1 \\ \text{over-reporting:} \quad & \phi_1 + 2\phi_2 > 1 \end{aligned}$$

- Notice that when $\phi_2 = 0$, the model results in simpler versions only accounting for under-reporting or no misreporting.

Method of moments

The marginal distribution of the observed process Y_n is essential to compute the moment-based estimates of the model:

$$Y_n = \begin{cases} \text{Poisson}(\mu_X) & 1-\omega \\ \text{2nd-order Hermite}(\mu_X\phi_2, \mu_X(1-\phi_1-\phi_2)) & \omega \end{cases} \tag{6}$$

1. Fit the mixture above to obtain estimates of ω, μ_X, ϕ_1 and ϕ_2
2. Use the theoretical expression of the ACF (ρ_Y) to estimate α
3. Using $\hat{\mu_X}$ (step 1) and $\hat{\alpha}$ (step 2), λ can be easily estimated

Maximum likelihood

Since the likelihood functions of the processes Y_n and Z_n are directly intractable (HMC with an infinite number of states), the forward algorithm is a reasonable choice to compute such functions. The likelihood is then computed by $P(Y_{1:N} = y_{1:N}) = \sum_{x_N=\frac{y_N}{2},1_N}^{\infty} \gamma_N(y_{1:N}, x_N, 1_N)$, where:

$$\gamma_n(y_{1:n}, x_n, 1_n) = \overbrace{\mathrm{P}(Y_n = y_n | X_n = x_n, 1_n)}^{\text{emission probabilities}} \sum_{x_{n-1}=\frac{y_{n-1}}{2},1_{n-1}}^{\infty} \underbrace{\mathrm{P}(X_n = x_n | X_{n-1} = x_{n-1})}_{\text{transition probabilities}}$$

$$\times \underbrace{\mathrm{P}(1_n = 1_n | 1_{n-1} = 1_{n-1})}_{\text{transition probability matrix}} \gamma_{n-1}(y_{1:n-1}, x_{n-1}, 1_{n-1}).$$

Forward probabilities

- While the transition probabilities are easily computed through the Poisson(λ)-INAR(1) cpmf, the emission probabilities are trickier:

model	$\mathrm{P}(Y_n = y_n \mid X_n = x_n, 1_n = 1_n)$
Y_n	$\begin{cases} 0 & \text{if} \quad x_n < \frac{y_n}{2} \\ 0 & \text{if} \quad 1_n = 0, x_n < \frac{y_n}{2} \\ p_n & \text{if} \quad 1_n = 1, x_n \geq \frac{y_n}{2} \\ 1 & \text{if} \quad 1_n = 0, x_n = y_n \end{cases}$

- p_n can be computed through the following recursive relation:

$$p_n = \frac{1}{n(1-\phi_1-\phi_2)}[\phi_1(x_n-(n-1))p_{n-1}+\phi_2(x_n-(n-2))p_{n-2}] \tag{7}$$

Simulation study

- In order to evaluate the model capabilities for under-reporting and over-reporting detection and estimation two time series are simulated, one with overreporting and another with under-reporting
- Although we here know which time series is over- and under-reported, the model also provides an easy mechanism for identifying which misreporting phenomenon is present in the data

Results

	Over-reporting				
	α	λ	ω	ϕ_1	ϕ_2
true parameter	0.3	3.0	0.7	0.1	0.8
point estimate	0.3578	3.2684	0.5501	0.0771	0.8072
std. error	0.1054	0.7352	0.1155	0.0297	0.0829

	Under-reporting				
	α	λ	ω	ϕ_1	ϕ_2
true parameter	0.5	3.0	0.7	0.2	0.1
point estimate	0.5184	3.0586	0.7354	0.1952	0.0784
std. error	0.1554	0.8808	0.0890	0.0511	0.0339

Results

We also compared several theoretical and empirical moments for both simulated time series, such as the mean, variance, and the first auto-correlation coefficients

- For the over-reported time series, we observed a mean and variance of 7.00 and 13.65, respectively, while the corresponding theoretical values are 6.91 and 14.31. With respect to the first auto-correlation coefficients, we observed $\hat{\rho}(1) = 0.192$, $\hat{\rho}(2) = 0.126$, and $\hat{\rho}(3) = 0.037$, while the corresponding theoretical values are $\rho(1) = 0.2183$, $\rho(2) = 0.0655$ and $\rho(3) = 0.0196$.

Results

- For the under-reported time series, the empirical mean and variance were 3.34 and 7.52 compared to the theoretical ones that were 3.40 and 6.82. The first three coefficients of the empirical auto-correlation were here $\hat{\rho}(1) = 0.163$, $\hat{\rho}(2) = 0.101$, and $\hat{\rho}(3) = 0.073$ compared to the theoretical ones that were $\rho(1) = 0.1433$, $\rho(2) = 0.0717$ and $\rho(3) = 0.0358$

Conclusions

- Several methodological approaches have been proposed recently to face underreported data based on count time series, but few can handle also overreported data
- The proposed model appropriately identifies whether the time series is over-reported or under-reported, and the true values of the parameters are always contained in the 90% Wald confidence intervals
- Among other applications, the proposed model can be applied to analyse the impact of the Covid-19 in health services usage

Amanda Fernández-Fontelo, Pedro Puig, Montserrat Guillén and David Moriña

Thank you!

dmorina@ub.edu