On Random Environment Integer-valued Autoregressive Models

Aleksandar Nastić

University of Niš, Faculty of Sciences and Mathematics, Višegradska 33, Niš
e-mail: aleksandar.nastic@pmf.edu.rs

Abstract. The main problem in modeling counting data is naturally defining the most suitable time series model, which would be the best possible representation of the observed real-life data. The data which we are interested are obtained by registering the correlated integer values of the particular phenomenon, or simply as a result of counting realizations of some phenomena, or even the elements of an observed population, in successive time intervals. The aforementioned data can be found in many fields of natural sciences and humanities including medicine, economics, finance, telecommunications, criminology, sports, etc. In order to model such data as well as possible, mathematicians used at first autoregressive time series with continuous marginal distributions. This gave acceptable results only when it comes to phenomena that generate extremely high realization values, where round-off error is negligible. On the other hand, in situations when observations registered over time are not that high (lower than 10^6), previously mentioned models cannot be used successfully. Slightly better results were achieved by involving Markov chains into the modeling procedure, as described in [1]. Unusually big number of parameters was the key obstacle to this idea. Few years later, several Discrete Autoregressive models of Moving Average ($DARMA$), based on well known $ARMA$ models, were defined in [2]. These models gave even better results. Finally, in mid 1980s, [3] and [4] introduced in different ways an Integer-valued Autoregressive model of order 1 ($INAR(1)$), based on the binomial thinning operator. This newly introduced model gave a foundation to the modern and contemporary approach in modeling the counting data sets.

However, INAR models based on binomial thinning were not suitable for describing, for example, population sizes through time in situations where it can change not only by their members disappearance, but also through their interactions or reproductions. With the aim od this kind model improvement, a significant breakthrough was made in [5] with the introduction of the negative binomial thinning operator and the so-called $NGINAR$ model with a geometric marginal distribution. After that, a large number of variations and generalizations of $INAR$ models, in terms of their orders, thinning operators, dimensionality and marginal distributions, have emerged in order to achieve more adequate fitting of counting data of various kinds.

All mentioned models were stationary. But very often certain non-stationary characteristics have appeared in the counting processes. So, that led us to our second milestone in INAR modeling, which was introducing a completely new concept of model construction by incorporating in data-fitting the effects of environment in which the counting process exists and which can change through time. We achieved this in [6] by introducing additional (controlling) process (Markov chain) which by taking different values of its random variables, so-called random states, in different periods of time, affects counting model values, by directly defining its marginal distribution parameter values. Later, in [7] and [8] some generalizations in respect of controlling the order of the model and correlation parameter value were presented using the same random state controlling process which was obtained by some process values clustering algorithms. Besides using well-known k-means clustering algorithm we have also tried to introduce more advanced clustering mechanism for our specific non-stationary INAR values, which was presented in [9] and [10]. Some bivariate random environment INAR models, also driven by a single process of random states, were successfully introduced in [11]

However, we have now realized that although the parameter values of the marginal distribution are intuitively driven by the random process defined by clustering the counting process values, that might not be the case with correlation parameter α. Namely, autocorrelation of the process should also depend on environment conditions, but not necessarily in the same way as marginals. Therefore, we introduce a new model with two control processes, where a separate and independent Markov process will determine the correlation in the model. It is also obtained by clustering, but not the values of the INAR process, but its sample autocorrelation functions. On this matter we have results given in [12].

Keywords: Random environment; INAR; RrNGINAR; negative binomial thinning; counting process.

References

[1] **D. R. Cox, H. D. Miller.** The theory of stochastic processes, *Methuen & Co*, London. 1965.

[2] **P. A. Jacobs, P. A. W. Lewis.** Discrete time series generated by mixtures I: correlational and runs properties, *Journal of the Royal Statistical Society, Series B*, 1978, 40(1), 94-105.

[3] **E. McKenzie.** Some simple models for discrete variate time series, *Journal of the American Water Resources Association*, 1985, 21(4), 645-650.

[4] **M. A. Al-Osh,A. A. Alzaid.** First-order integer-valued autoregressive (INAR(1)) process, *Journal of Time Series Analysis*, 1987, 8(3), 261-275.

[5] **M. M. Ristić, H. S. Bakouch, A. S. Nastić.** A new geometric first-order integer-valued autoregressive (NG-INAR(1)) process, *Journal of Statistical Planning and Inference*, 2009, 139(7), 2218-2226.

[6] **A. S. Nastić, P. N. Laketa, M. M. Ristić.** Random environment integer-valued autoregressive process, *Journal of Time Series Analysis*, 2016, 37(2), 267-287.

[7] **A. S. Nastić, P. N. Laketa, M. M. Ristić.** Random environment INAR models of higher order, *REVSTAT-Statistical Journal*, 2019, 17(1), 35-65.

[8] **P. N. Laketa, A. S. Nastić, M. M. Ristić.** Generalized random environment INAR models of higher order, *Mediterranean Journal of Mathematics*, 2018, 15(1):9.

[9] **B. A. Pirković, P. N. Laketa, A. S. Nastić.** On generalized random environment INAR models of higher order: Estimation of random environment states, *Filomat*, 2021, 35(13), 4545-4576.

[10] **B. A. Pirković, M. M. Ristić, A. S. Nastić.** Random Environment Integer-Valued Autoregressive Process with Discrete Laplace Marginal Distributions, *REVSTAT-Statistical Journal*, 2023, 21(4), 469-490.

[11] **P. M. Popović, P. N. Laketa, A. S. Nastić.** Forecasting with two generalized integer-valued autoregressive processes of order one in the mutual random environment, *SORT Statistics and Operations Research Transactions*, 2019, 43(2), 355-384.

[12] **A. S. Nastić, M. S. Djordjević, T. D. Čamagić** Value-Correlation Controlled Integer-valued Autoregressive Process in Random Environment, *submitted*, 2024.