DOI: 10.1478/AAPP.972A2

AAPP | Atti della Accademia Peloritana dei Pericolanti
Classe di Scienze Fisiche, Matematiche e Naturali
ISSN 1825-1242

Vol. 97, No. 2, A2 (2019)

MODELLING AND FORECASTING TEMPERATURE AND PRECIPITATION IN ITALY

MARIO LEFEBVRE *

ABSTRACT. We study the monthly average temperature in Italy for the period 1991-2015. The increase or decrease of the average temperature with respect to the previous year is modelled as a discrete-time Markov chain having four possible states. Similarly, a Markov chain is proposed as a model for the variations of the monthly amount of precipitation. Based on these models, it is possible to forecast whether the temperature and the amount of precipitation are likely to vary significantly in the long term.

1. Introduction

Stochastic processes, and in particular Markov chains, have been used by many authors to model various phenomena. For example, Avilés *et al.* (2016) forecast drought events based on Markov chains. Drton *et al.* (2003) proposed a Markov chain model of tornadic activity. Matis *et al.* (1989) used these processes to forecast cotton yields. Tagliaferri *et al.* (2016) generated artificial wind speed time series with the help of nested Markov chains. In hydrology and meteorology, Markov processes are very useful to forecast river flows and precipitation (see, for instance, Caskey Jr. 1963; Lefebvre and Guilbault 2008).

Our objective in this paper is first to find a stochastic model for the variations of average monthly temperature and rainfall in the case of Italy, and then to use this model to forecast their long-term evolution. A discrete-time Markov chain having four possible states will be proposed as a model for temperature and also for precipitation.

The World Bank has created the Climate Change Knowledge Portal. On the website http://sdwebx.worldbank.org/climateportal, one can find the historical average monthly temperatures and rainfalls for all the countries in the world. The data are available for the period 1901-2015. We will use the data for the years 1991-2015 to estimate the various transition probabilities of the Markov chains. The limiting probabilities of the Markov chains will then be computed to try to forecast the long-term behaviour of the variables of interest. The data set will also be divided into two parts to check whether there have been some significant changes in the variations of temperature and rainfall during the period considered. In the next section, we will present the model that we propose, and then the model will be implemented for temperature and rainfall in Sections 3 and 4, respectively.

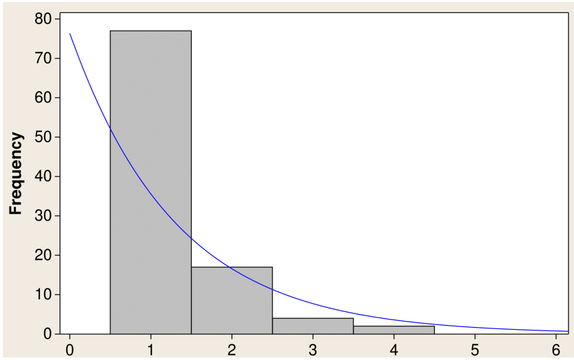

FIGURE 1. Histogram of a geometric distribution.

2. Discrete-time Markov chains

Let $X_0, X_1, \ldots$ be random variables. Suppose that the possible values of the random variables are the integers $S := \{0, 1, \ldots\}$ (where the symbol ":=" means *by definition*). The stochastic process $\{X_n, n = 0, 1, \ldots\}$ is a (discrete-time) Markov chain if

$$P[X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0] = P[X_{n+1} = j \mid X_n = i] \tag{1}$$

for all states $i_0, \ldots, i_{n-1}, i, j$ in S, and for any $n \in \{0, 1, \ldots\}$. That is, the *conditional probability* that the process will be in state j at time $n+1$, *given* the whole history of the process since the initial state i_0 actually only depends on the most recent value of the state. We say that the future $(n+1)$, given the present (n) and the past $(n-1, \ldots, 0)$, depends only on the present. In general, we also assume that

$$p_{i,j} := P[X_{n+1} = j \mid X_n = i] = P[X_1 = j \mid X_0 = i] \tag{2}$$

for any $n \in \{0, 1, \ldots\}$. That is, we assume that the Markov chain is time-homogeneous, because the conditional probability does not depend on the time n. The $p_{i,j}$, for $i, j \in S$, are called the one-step transition probabilities of the Markov chain, and the matrix $\mathbf{P} := (p_{i,j})_{i,j \in S}$ is the transition matrix.

Let T_i be the number of time units that the process spends in state i before making a transition to another state. By independence, we have that

$$P[T_i = k] = p_{i,i}^{k-1}(1 - p_{i,i}), \quad \text{for } k = 1, 2, \ldots . \tag{3}$$

That is, T_i has a *geometric distribution* with parameter $p := 1 - p_{i,i}$.

An example of a histogram obtained by generating 100 observations from a geometric distribution with parameter $p = 0{,}8$ is shown in Fig. 1, in which an exponentially decreasing function has been added to show the expected form of the histogram.

Now, under some conditions (that will be fulfilled in the applications considered in this paper), we can show that the limiting probabilities

$$\pi_i := \lim_{n \to \infty} P[X_n = i] \tag{4}$$

exist and can be obtained by solving the system (see, for example, Lefebvre 2007):

$$\pi = \pi \mathbf{P}, \tag{5}$$

subject to

$$\sum_{i=0}^{\infty} \pi_i = 1, \tag{6}$$

where

$$\pi := (\pi_0, \pi_1, \ldots). \tag{7}$$

These limiting probabilities, which also represent the proportion of time that the process spends in state i over a long period, when it is in equilibrium, will enable us to forecast the long-term behaviour of the corresponding Markov chain. Indeed, if the model proposed for the monthly average temperatures and the monthly amounts of precipitation in Italy is deemed to be realistic, then by computing the limiting probabilities π_i we can determine what should happen after a large number of months when the process stabilizes. Of course, the forecasts will be based on the observations in the data set that were used to compute these limiting probabilities. The model should be updated and the π_i recalculated when new observations become available. In the next section, the above model will be implemented in the case of average monthly temperatures in Italy.

3. Average monthly temperature

Let D_n be the difference between the average temperature (in degrees Celsius) for a given month and the corresponding month of the previous year. We define

$$X_n = \begin{cases} 0 & \text{if } D_n \leq -1, \\ 1 & \text{if } D_n \in (-1, 0), \\ 2 & \text{if } D_n \in [0, 1), \\ 3 & \text{if } D_n \geq 1. \end{cases} \tag{8}$$

First, using the data from http://sdwebx.worldbank.org/climateportal for the period 1991-2015, we obtained the histograms of the random variables T_i, for $i = 0, 1, 2, 3$. They are shown in Figs. 2-5. We see that the various histograms have indeed the form expected for a geometric distribution. Therefore, a Markov chain with the above states is a realistic model for the *variations* of the monthly average temperature in Italy during the period considered, which implies that the limiting probabilities π_i can be used to forecast the variations of the monthly average temperatures after a large number of months.

Remark. We could perform Pearson's goodness-of-fit statistical test to check whether the geometric distribution is a good model for the random variables T_i. However, here the number of *degrees of freedom* would be very small.

Next, we estimated the transition probabilities $p_{i,j}$. The estimated transition matrix is

$$\mathbf{P} = \begin{bmatrix} 5/67 & 12/67 & 11/67 & 39/67 \\ 12/69 & 16/69 & 24/69 & 17/69 \\ 16/64 & 20/64 & 16/64 & 12/64 \\ 37/76 & 22/76 & 11/76 & 6/76 \end{bmatrix}. \tag{9}$$

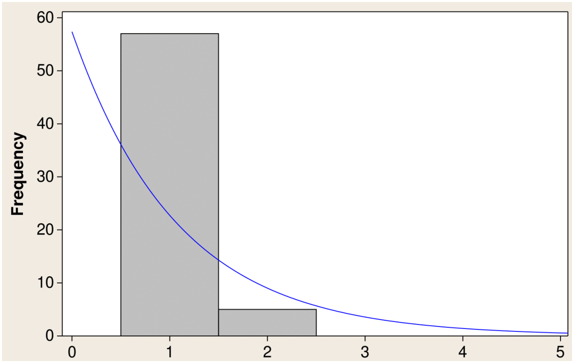

FIGURE 2. Histogram of the random variable T_0 in the case of temperature.

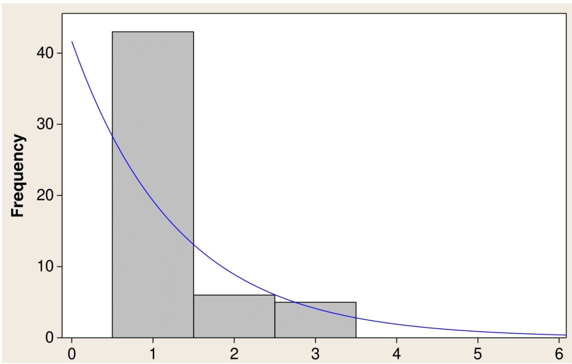

FIGURE 3. Histogram of the random variable T_1 in the case of temperature.

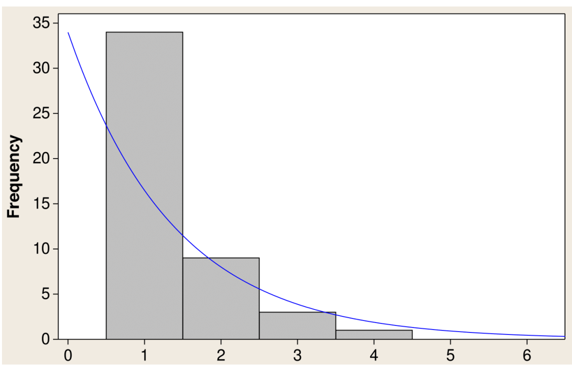

FIGURE 4. Histogram of the random variable T_2 in the case of temperature.

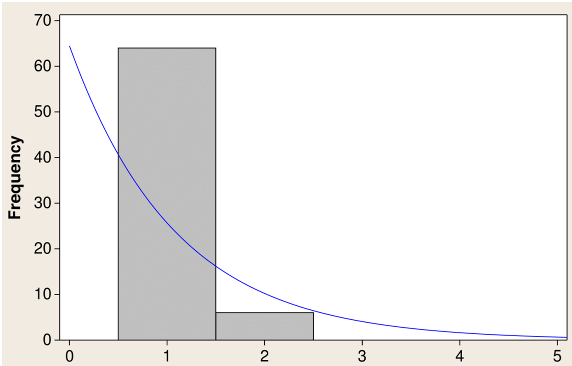

FIGURE 5. Histogram of the random variable T_3 in the case of temperature.

Since the elements of the matrix are all different from zero, it follows immediately that all the conditions needed for the existence of the limiting probabilities are fulfilled. These limiting probabilities can therefore be obtained by solving (5), (6). We find that

$$\pi_0 = 0{,}2512, \quad \pi_1 = 0{,}2524, \quad \pi_2 = 0{,}2245, \quad \pi_3 = 0{,}2719. \tag{10}$$

Notice that the four limiting probabilities are not very different, with a slightly larger probability of having an increase of at least 1 degree. The probability of having a decrease (0,5036) is almost the same as that of having an increase (0,4964) of the average monthly temperature. Actually, the mean of the 288 monthly differences is approximately 0,065, which is rather small.

Remark. By definition, the π_i are *limiting* probabilities. They can also be obtained by computing $\mathbf{P}^n$. For n large enough, the four lines of the matrix $\mathbf{P}^n$ should be almost identical. Here, we find that

$$\mathbf{P}^{16} \simeq \begin{bmatrix} 0{,}2512 & 0{,}2524 & 0{,}2245 & 0{,}2719 \\ 0{,}2512 & 0{,}2524 & 0{,}2245 & 0{,}2719 \\ 0{,}2512 & 0{,}2524 & 0{,}2245 & 0{,}2719 \\ 0{,}2512 & 0{,}2524 & 0{,}2245 & 0{,}2719 \end{bmatrix}. \tag{11}$$

Therefore, after 16 months the probabilities of the four possible states of the Markov chain are already equal to the limiting probabilities.

To check whether the conclusions are the same if we consider only the most recent years, we divided the data set into two equal parts (from the beginning of 1991 to June 2003, and from July 2003 to the end of 2015), and we computed the limiting probabilities for each part. The results are shown in Table 1.

We notice that π_2 and π_3 did not change much between the beginning and the end of the time period considered, but that the limiting probabilities π_0 and π_1 have been almost inverted, denoting a larger probability of big decreases. The probability of large positive or negative changes has increased from about 0,49 to more than 0,55. Hence, one may conclude that climate change seems to be observable in Italy. In the next section, we will turn to average monthly rainfalls in Italy.

TABLE 1. Limiting probabilities for the time periods 1991-2015, 1991-2003 and 2003-2015 in the case of temperature.

Period	π_0	π_1	π_2	π_3
1991-2015	0,2512	0,2524	0,2245	0,2719
1991-2003	0,2227	0,2884	0,2171	0,2718
2003-2015	0,2821	0,2155	0,2292	0,2732

4. Average monthly rainfall

Let now D_n denote the difference between the average rainfall (in mm) for a given month and the corresponding month of the previous year. We set

$$X_n = \begin{cases} 0 & \text{if } D_n \leq -30, \\ 1 & \text{if } D_n \in (-30, 0), \\ 2 & \text{if } D_n \in [0, 30), \\ 3 & \text{if } D_n \geq 30. \end{cases} \tag{12}$$

As for the case of average monthly temperature, we first look at the histograms of the random variables T_i, for $i = 0, 1, 2, 3$; see Figs. 6-9. Again, we can observe that the various histograms seem to confirm the fact that the corresponding random variables have approximately a geometric distribution, as is required for the model to be realistic. The estimated transition probabilities $p_{i,j}$ are the following:

$$\mathbf{P} = \begin{bmatrix} 1/59 & 12/59 & 19/59 & 27/59 \\ 14/79 & 19/79 & 31/79 & 15/79 \\ 19/84 & 30/84 & 28/84 & 7/84 \\ 25/54 & 19/54 & 7/54 & 3/54 \end{bmatrix}, \tag{13}$$

and the limiting probabilities are given by

$$\pi_0 = 0{,}2122, \quad \pi_1 = 0{,}2897, \quad \pi_2 = 0{,}3097, \quad \pi_3 = 0{,}1884. \tag{14}$$

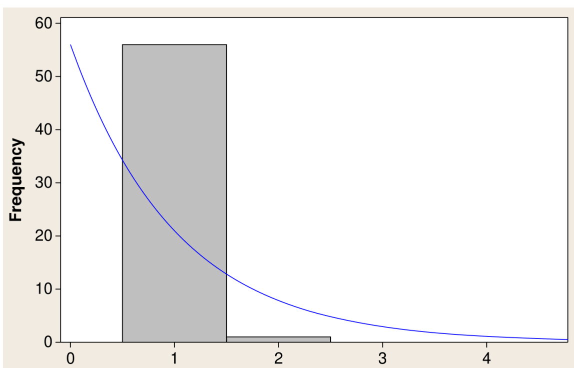

FIGURE 6. Histogram of the random variable T_0 in the case of precipitation.

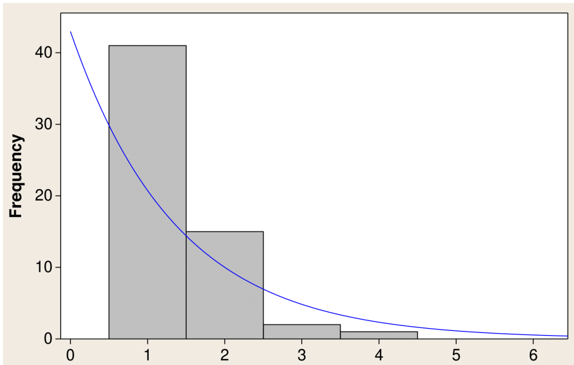

FIGURE 7. Histogram of the random variable T_1 in the case of precipitation.

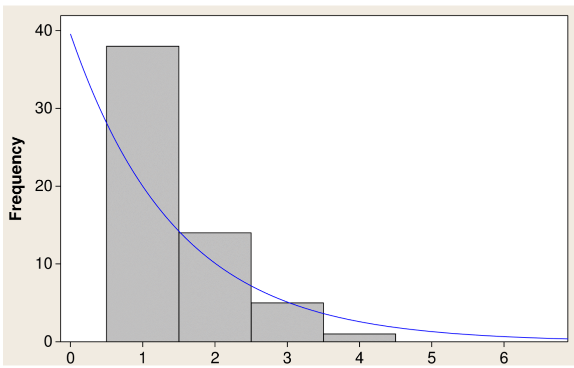

FIGURE 8. Histogram of the random variable T_2 in the case of precipitation.

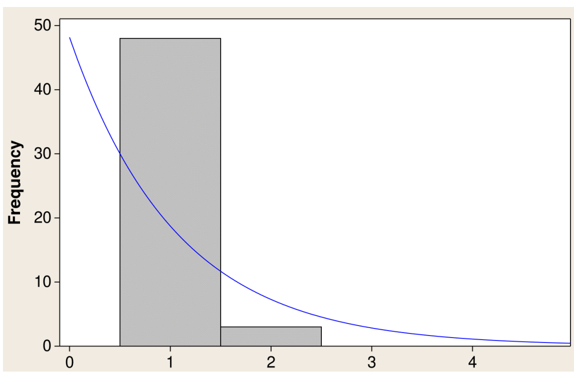

FIGURE 9. Histogram of the random variable T_3 in the case of precipitation.

Remark. As in the case of temperature, we find that the four rows of the matrix $\mathbf{P}^{16}$ are equal to the above probabilities. It follows that the forecasts that we make are valid after at most 16 months.

This time, we have much more variation in the limiting probabilities and a smaller probability (0,4006) of large changes (with the choice of 30 mm in the definition of the various states). However, the probability of having a decrease (0,5019) is again almost the same as that of having an increase (0,4981) of the average monthly rainfall. The mean of the 288 monthly differences is approximately $-0{,}36$ mm, pointing indeed toward a small decrease in the long run.

Finally, we performed the same analysis with the two subsets of data of the same size. The results, presented in Table 2, are quite similar, except for a large increase in the value of π_3. The values of π_0 and π_1 are rather constant, but π_2 has decreased and π_3 has increased

TABLE 2. Limiting probabilities for the time periods 1991-2015, 1991-2003 and 2003-2015 in the case of rainfall.

Period	π_0	π_1	π_2	π_3
1991-2015	0,2122	0,2897	0,3097	0,1884
1991-2003	0,2088	0,2965	0,3278	0,1668
2003-2015	0,2155	0,2829	0,2914	0,2102

significantly. Also, there has been an increase from a 0,38 probability of large changes to 0,43. A regime change seems to have taken place if we look at the means of the 144 monthly differences of each subset: respectively $+0{,}11$ and $-0{,}83$. Climate change therefore creates even more variations in rainfall than in temperature.

5. Conclusion

In this paper, we studied the average monthly temperatures and rainfalls in Italy during the period 1991-2015. We saw that, in each case, a discrete-time Markov chain having four possible states is an appropriate model for the data. Based on our models, we were able to estimate the long-term behaviour of the changes in temperature and amount of precipitation. The data seem to confirm the effects of climate change on both temperature and rainfall in Italy, rainfall being especially affected.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada. The author also wishes to thank the reviewer for his/her constructive comments.

References

Avilés, A., Célleri, R., Solera, A., and Paredes, J. (2016). "Probabilistic forecasting of drought events using Markov chain- and Bayesian network-based models: A case study of an Andean regulated river basin". *Water* **8**(2), 37. DOI: 10.3390/w8020037.

Caskey Jr., J. E. (1963). "A Markov chain model for the probability of precipitation occurrence in intervals of various length". *Monthly Weather Review* **91**(6), 298–301. DOI: 10.1175/1520-0493(1963)091%3c0298:AMCMFT%3e2.3.CO;2.

Drton, M., Marzban, C., Guttorp, P., and Schaefer, J. T. (2003). "A Markov chain model of tornadic activity". *Monthly Weather Review* **131**(12), 2941–2953. DOI: 10.1175/1520-0493(2003)131%3c2941:AMCMOT%3e2.0.CO;2.

Lefebvre, M. (2007). *Applied Stochastic Processes*. New York, United States: Springer-Verlag. DOI: 10.1007/978-0-387-48976-6.

Lefebvre, M. and Guilbault, J.-L. (2008). "Using filtered Poisson processes to model a river flow". *Applied Mathematical Modelling* **32**(12), 2792–2805. DOI: 10.1016/j.apm.2007.09.035.

Matis, J. H., Birkett, T., and Boudreaux, D. (1989). "An application of the Markov chain approach to forecasting cotton yields from surveys". *Agricultural Systems* **29**(4), 357–370. DOI: 10.1016/0308-521X(89)90097-8.

Tagliaferri, F., Hayes, B. P., Viola, I. M., and Djokić, S. Z. (2016). "Wind modelling with nested Markov chains". *Journal of Wind Engineering and Industrial Aerodynamics* **157**, 118–124. DOI: 10.1016/j.jweia.2016.08.009.

* Polytechnique Montréal
Department of Mathematics and Industrial Engineering
C.P. 6079, Succursale Centre-ville, Montréal, Québec, Canada H3C 3A7

Email: mario.lefebvre@polymtl.ca

Communicated 6 December 2018; manuscript received 7 January 2019; published online 6 July 2019