

Model fitting for predicted precipitation in Darwin: Some issues with model choice

Jim Farmer

Macquarie University
<jim.farmer@mq.edu.au>

In Volume 23(2) of the *Australian Senior Mathematics Journal*, Boncek and Harden present an exercise in fitting a Markov chain model to rainfall data for Darwin Airport (Boncek & Harden, 2009).

Days are subdivided into those with precipitation and precipitation-free days. I will abbreviate these labels to wet days and dry days. It is suggested that a 2 state Markov chain model may be suitable for modelling the pattern of wet and dry days.

1. As a first attempt, the data for calendar year 2008 are used to fit the transition probabilities. The model is tested by using the stationary distribution of the Markov chain to predict the number of wet and dry days in the period 1999–2008. A chi-squared test is used to compare the predicted numbers with the actual numbers and this test suggests the model is not reliable.
2. The data are examined in more detail and it is found that 2008 was an unusually dry year. As a second attempt, the transition probabilities are refitted using the two years' data 2007–2008. The numbers of wet and dry days for 1999–2008 are again predicted and compared to actual data via the chi-squared test, this time finding no significant variation.

The conclusion made is that for the second attempt, “this forecast model works” (Boncek & Harden, 2009, p. 14), while the first attempt did not work.

Relevance to Australian High School Curriculum

Markov chains appear in the Mathematical Methods (CAS) subject and the Specialist Mathematics subject in Victoria (VCAA, 2010), the Mathematics C subject in Queensland (QSA, 2008) and are mentioned in a draft future syllabus for the Mathematical Methods subject in South Australia (SACE Board of SA, 2010). Markov chains do not currently appear in the draft Australian Senior Secondary Curriculum (ACARA, 2009). Where they do appear in state syllabi, generally, they are introduced as an example of an application of matrices without discussion of methods of assessing the good-

ness of fit of the model. Hence, in the Australian context the type of modelling exercise described by Boncek and Harden is probably more suited to the tertiary education sector.

Use of the chi-squared test

The chi-squared test employed by Boncek and Harden is usually encountered when dealing with a multinomial random variable. For example, there may be n independent trials each of which can result in one of r outcomes with $r \geq 3$, and we are attempting to verify whether the number of each outcome observed from n trials is consistent with a model that proposes the probability of each outcome occurring. The relevant chi-squared random variable has $r - 1$ degrees of freedom, which is at least two.

If there are only two possible outcomes for each trial, a chi-squared test with only one degree of freedom is still technically valid, but is unnecessarily complex. The underlying random variable is now simply binomial, so there is a simpler way to test the hypothesis.

Considering Table 8 of Boncek and Harden's paper (2009, p. 13), the hypothesis is that the number of dry days has the binomial distribution $\text{Bi}(3653, 66.58\%)$ and we are testing whether an observation of 2416 dry days is consistent with that. Employing a hypothesis test with the binomial random variable—or by employing the Central Limit Theorem to approximate it by a normal random variable—might bring the example within the range of understanding of more students than does use of the chi-squared test.

To view this issue from a different perspective, a chi-squared random variable with n degrees of freedom arises from summing the squares of n independent normal random variables. Hence, a chi-squared random variable with 1 degree of freedom is simply the square of a normal random variable. As a rule of thumb, if you find yourself using a chi-squared test with only one degree of freedom, consider whether there is a simpler way to view the problem.

Checking the hypothesis with overlapping data

The type of hypothesis test employed to test the goodness of fit of the model assumes the data being used to test the model is independent of the data used to fit the model. Here the model was first fitted with 2008 data and then tested against a 10-year set of data that included 2008. It was then refitted using 2007 and 2008 data and tested against a 10-year set of data that included both years. We would naturally expect the second model to more closely fit the 10 years of data since there is greater overlap, but there is nothing in the hypothesis testing process that demands a higher degree of coherence in the results before classifying the model as acceptable.

The Markov chain model is complex, so perhaps a simpler model might clarify the problem here. If we were to estimate the number of dry days in the

10 years 1999–2008 by using p_1 , defined as the proportion of days in 2008 that were dry, we might expect the estimate to be moderately good. If we do it using p_2 , the proportion of days in 2007–2008 that were dry, we would expect a better estimate. However, if we estimate the number of dry days in the 10 years 1999–2008 by using p_{10} , the proportion of days in that 10 years that were dry, we get exactly the right answer. It would then *not* be appropriate to test the fit of the model by using a hypothesis test that assumes the number of dry days in 1999–2008 is a binomial random variable with distribution $\text{Bi}(3653, p_{10})$, since that number is not a random variable at all. It is a fixed number that was used to fit the parameter p_{10} .

If however we were suggesting p_{10} was useful for wider time periods, we could test this by using data from a different set of 10 years, such as 1989–1998.

Returning to the proposed Markov chain model, to test whether the model fitted using 2007–2008 data is appropriate, we would need to use data that includes neither 2007 nor 2008.

Inappropriateness of the Markov chain model

Markov chains as considered here have the time-homogeneous property, meaning that the transition probabilities are constant over time. It is possible to develop Markov chains that do not have this property but they have fewer interesting mathematical results. Most of the nice results about Markov chains require the time homogeneous property, so usually an unqualified reference to “Markov chains” means those that have the time-homogeneous property. (For similar reasons, it usually also means they have a finite number of states.)

In their “What went wrong?” section, Boncek and Harden suggest that their first attempt at fitting the model failed because the transition probabilities were not constant, with 2008 being a particularly dry year. They then refit the model using 2007–2008 data, giving parameters that more closely match the 10 year average and conclude this Markov chain model is appropriate. This misses the major reason that the Markov chain model is not appropriate, which is that the transition probabilities vary enormously *within* each year.

Darwin, being in the tropics, has two seasons, wet and dry. From May to October, if it was dry today it will almost certainly be dry tomorrow, because almost all days in this range are dry. For 2008 there were 178 dry days in these months and in 175 cases they were followed by a dry day. From January to March, if it was dry today there is a significant chance it will be wet tomorrow, because there are many wet days from January to March. For 2008 there were 23 dry days in this period and 13 of these were followed by a wet day.

Another way of viewing this problem is to imagine using the fitted Markov chain to run a simulation of wet and dry days for the next 10 years. Within each calendar year the simulation would tend to spread the wet days evenly across the year, where as in practice the wet days should be clumped in the wet season.

What really went wrong?

There are two possible perspectives here, and I am unsure which the authors intended.

The first perspective is: “We’re teaching Markov chains. Let’s pick an interesting set of data, fit a Markov chain and then talk about how to test whether it’s a suitable model.”

In this perspective, the test Boncek and Harden apply is not sufficiently specific. The Markov chain model makes predictions about the pattern of wet and dry days. It says they are uniform across the year. The test only looked at the total number of dry days for a certain number of complete years. Even if the model correctly predicts the number of dry days over 10 years, it is still an inappropriate model since it doesn’t correctly predict the clumping of wet days in the wet season.

No detailed calculations are required to reject the model. A time line showing the actual and predicted pattern of dry days over various calendar years will suffice to demonstrate that the model is not capturing the seasonal pattern of Darwin rainfall.

The second perspective is: “The aim is to predict the number of dry days at Darwin airport next year. Would a Markov chain model give a good answer?”

In this situation some people may argue that a model that incorrectly predicts the distribution of dry days across the year might still be useful if it accurately models the total number of dry days in the year. I find this unconvincing. Ockham’s razor seems relevant here. If the aim is only to estimate the total number of dry days in a year, the logical starting point is to average the number of dry days in the years from recent history. There seems no obvious reason to build a model that predicts the pattern of dry days within the year if we only seek to estimate the total number of dry days.

However, if the task is to estimate the *distribution* of the number of dry days that will occur next year, a more detailed model may be appropriate. For example, a first glance at the data suggests that the arrival of the wet season might vary by a week or so from year to year, so it may be useful to build a model of the start date (and end date) of the wet season.

What type of calculations could justify a Markov chain model?

Continuing the above idea, might it be possible to use different Markov chain models with different transition probabilities for different seasons? To take a simpler problem: January always falls entirely within the wet season. Could we model the precipitation status of days in January by a Markov chain?

Appealing to Ockham’s razor, to model the number of dry days in January, the starting point would be a binomial model that has each day in January having a probability p of being dry, with the result of each day being independent of every other day. By contrast, to justify a more complex model such

as a Markov chain, we would need some evidence of serial correlation; that is, we would want some evidence that the probability of tomorrow being dry is influenced by whether today is dry.

Let p_D be the probability it will be dry tomorrow if it was dry today.

Let p_W be the probability it will be dry tomorrow if it was wet today.

We adopt a null hypothesis that $p_D = p_W$, which would justify the independence model, and seek statistical evidence that $p_D \neq p_W$, which would give grounds for further investigation of a Markov chain model.

Consider January 2008, it had 13 dry days, eight of which were followed by a dry day ($p_D = 62\%$) and had 17 wet days, six of which were followed by a dry day ($p_W = 35\%$). These two percentages look quite different, but a 2 proportion test finds them not statistically significant at the 5% level. So, as yet there is no evidence to implement a Markov model rather than simply assuming the precipitation status on each day in January is independent of every other day.

This is of course a cop out! Since the sample size is small, if the independence assumption is true we can still easily get very large differences between the observed values of p_D and p_W , so there would need to be very large differences between their true values for the test to fail. What we should do is increase the sample size by collating January data for perhaps 10 years. The extraction of data from the relevant web pages is quite tedious, so this is left as an exercise for readers more skilled at data manipulation than I.

Existing literature

Thanks go to an anonymous reviewer and the editors for supplying pointers to some existing literature on the topic of fitting Markov chain models to rainfall data. The idea of using a Markov chain for modelling wet and dry days seems to be due to Gabriel and Neumann (1962). It is instructive to contrast their approach to that of Boncek and Harden, and to my comments above.

Gabriel and Neumann use daily rainfall data for Tel Aviv, classifying each day as wet or dry. A rainfall reading of 0.1mm or more results in a day being classified as wet. They recognise that Tel Aviv, like Darwin, has a rainfall pattern that varies over the year. They only investigated data for the rainy season, running from November to April. Their data covers the rainy seasons from 1923/24 to 1949/50. Even within the rainy season, they recognise the true probabilities or rain vary within each calendar month. However, they try fitting a model which assumes that the relevant probabilities vary by calendar month but are constant within each month and find this gives a good fit. They note that the fitted probabilities do not vary greatly within the mid-winter months of December to February and further investigation leads them to conclude that fitting a constant set of transition probabilities to the whole mid-winter period still produces a reliable model.

I suggested above that the first step in testing whether a Markov chain model is suitable for modelling the pattern of wet and dry January days in Darwin would be to test whether we can justify the simpler model of occurrence of rain on successive days being independent. This type of investigation

may be within the abilities of many tertiary students.

However, with their greater knowledge of the existing literature, Gabriel and Neumann began with the knowledge that the simpler model was unlikely to succeed. They cite earlier studies showing that in some locations the probability that a wet day will be followed by another depends on the current number of consecutive wet days experienced. For them the question is not whether a model simpler than a Markov chain will provide a good fit, but rather whether Tel Aviv may be a location where something as simple as a Markov chain will give a good fit, when such a model is clearly not appropriate in other locations. Hence they employ a different method of testing their model's goodness of fit. For their primary test:

“The fit of the Markov chain model is examined by testing whether the proportions of wet days, given the previous day's weather, are independent of the weather two or more days earlier” (p. 93).

While the interested reader can find the details in Gabriel and Neumann's paper, the complexity of the test probably makes it unsuitable as a classroom exercise other than for statistics majors.

It has been subsequently shown by Green (1964) that more complex models can give a better fit to the Tel Aviv rainfall data. However, Gabriel and Neumann wrote at a time when computers were not readily available. Hence, knowing that the relatively simple Markov chain model gave an acceptable fit was of great value, even if there were more complex models that could give a better fit. It is interesting to note that Gabriel and Neumann explicitly reference a table of logarithms of binomial coefficients which they describe as “indispensable” to their work!

Disclaimer

The above arguments should be read with the disclaimer that I have no skills in meteorology. The problem has been approached solely as an exercise in fitting a statistical model to data, with no understanding of the mechanics behind the data.

Modelling can benefit enormously from experts in the field. A skilled meteorologist might be able to supply good physical reasons for adopting a model that might not be immediately obvious to us from a mere 10 years' data. For example, there could conceivably be a physical reason for the very rare rain in the dry season to mostly happen as midnight rainstorms that cause rain to be registered on two consecutive days. This would suggest the need for a model more complex than the Markov chain model. Alternatively, depending on the intended use of the model, it might be decided that wet days in the dry season are so rare as to be not worth a complex model.

References

- Australian Curriculum Assessment and Reporting Authority (ACARA). (2009). *Draft senior years curriculum: Mathematics*. Available from www.australiancurriculum.edu.au
- Boncek, J., & Harden, S. (2009). Predicting precipitation in Darwin: An experiment with Markov chains. *Australian Senior Mathematics Journal*, 23(2), 8–14.
- Gabriel, K. R., & Neumann, J. (1962). A Markov chain model for daily rainfall occurrence at Tel Aviv. *Quarterly Journal of the Royal Meteorological Society*, 88(375), 90–95.
- Green, J. R. (1964). A model for rainfall occurrence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 345–353.
- Queensland Studies Authority (QSA). (2008). *Senior syllabus: Mathematics C*. Brisbane: QSA.
- South Australian Certificate of Education (SACE) Board of South Australia (2010). *Mathematical Methods Draft 2011 Subject Outline: Stage 2*. Adelaide: SACE Board of SA. Available from <http://www.sace.sa.edu.au/subjects/stage-2-in-2011/mathematics/mathematical-methods>
- Victorian Curriculum and Assessment Authority. (2010). *Mathematics Victorian Certificate of Education Study Design*. Melbourne: VCAA. Available from <http://www.vcaa.vic.edu.au/vce/studies/mathematics/mathstd.pdf>.