SOCIETY OF ACTUARIES

Article from:

# Forecasting & Futurism

July 2013 – Issue 7

# **Hidden Markov Models** and You

*By Doug Norris & Brian Grossmiller*

This is the first of a series of articles exploring uses of Hidden Markov Models in actuarial applications. In this introduction, we will go over the basics of Hidden Markov Models along with some brief illustrative examples.

## WHAT IS A HIDDEN MARKOV MODEL?

A Hidden Markov Model (HMM) is a method for evaluating, and finding patterns within, time series. This can be a very useful way to model data under the right circumstances, such as when individual data points could be swayed by different influences. For instance, the data in Figure 1 shows some clear low and high periods, indicating that some manner of mixture model could be a good fit.

**Figure 1**: Example of data showing high and low patterns.

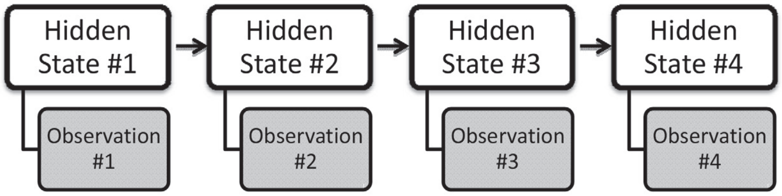
Some interesting questions arise from patterns like these, particularly when we see them in claims or sales volume data. How do we know if we are in a high or low state? What are the odds of staying high or low? What do we think the next data point will look like?

HMMs are similar to the Markov Processes covered in the current actuarial syllabus. The critical difference is that the matrix of state transitions is hidden, and we have to infer it from our data. Fortunately, the transition matrix typically is quickly estimated using a computer; we'll get into a basic example after we review some assumptions and definitions.

## STARTING WITH THE BASICS

When our data contains categorical information, we can construct a sequence of states easily. In the absence of defined categories, the real power of HMMs comes into play, because we have to use the data values themselves to determine these states or categories. A set of observations associated with the sequence of hidden states can infer the (hidden) transition matrix. A basic illustration of the hidden states and resulting observations is shown in Figure 2.

**Figure 2**: Hidden States and Associated Observations




In order to apply an HMM, we need to assume that each observation is drawn from a given distribution. Discrete distributions, such as the Poisson, Binomial, or Negative Binomial, are attractive due to their mathematical tractability. Using a computer, several distributions can be constructed rapidly, along with statistics measuring how well they fit. In practice it may prove convenient to model them all and select a distribution on the back end.

In addition to the matrix of transition probabilities between states, the data also can determine the parameters for the assumed probability distributions in each state. We'll illustrate this with an example later in the article, but for now we'll go

CONTINUED ON **PAGE 18**

over some of the benefits of using an HMM in the context of our three earlier questions.

### What state are we in?

Decoding is the process of using our data to determine the most likely sequence of hidden states we have observed. When we are looking at claims or sales volume, we will be interested in the state we are observing at each point (known as local decoding).

Conversely, in applications where the hidden state is an in-or-out type category (such as the presence or absence of a chronic disease), we are more interested in what state all of the observations collectively imply. This is known as global decoding and is the method used in many applications of speech and facial recognition technology.

### What state will we be in?

Our decoded data can give us a useful estimate of what the current state might be, and combined with the transition matrix we can estimate what state we are likely to be in over the near term. If we estimate a large probability of transitioning into a different state at the next time point then the most recent data may be a poor estimator.

In the forthcoming example, we will assume that the transition at any point in time only depends upon the state immediately preceding it. In practice, you might see higher order dependencies. For instance, visual inspection may suggest that your data shows a "high" state lasting for about three time periods before transitioning. In this case, modeling the transitions to depend upon the past three states would make intuitive sense.

THE EM ALGORITHM IS AN ITERATIVE METHOD, WHICH PRODUCES MAXIMUM LIKLIHOOD ESTIMATES FOR MISSING PARAMETERS IN AN HMM MODEL SOLUTION.

### What do we think the next data point will look like?

In addition to the transition matrix, constructing an HMM also requires estimating the probability distribution in each state. If we have assumed a convenient distribution, then calculating the expected values of each state is straightforward. Our estimate of the probabilities of future states can serve as weights to determine the value we expect to see at that point in time.

Now, let's take a look at an actual example to get an idea of how HMMs work in action.

## ESTIMATING HIDDEN MARKOV MODEL SOLUTIONS

Because Hidden Markov model solutions are not easily solved in a closed form, many numerical analysis algorithms have been built to estimate HMM parameters. The most popular of these algorithms is known as the EM algorithm (or the Baum-Welch algorithm, after its inventors).

The EM algorithm is an iterative method, which produces maximum likelihood estimates for missing parameters in a HMM model solution. This algorithm involves repeatedly invoking an "E step" (estimating the conditional expectation of the functions generated from the missing data) and an "M step" (maximizing the likelihood, where the functions are replaced by those conditional expectations), until the algorithm converges upon a solution. The solution space is typically non-linear, and will depend upon our initial estimate of the HMM parameters.

These algorithms are programmatically straightforward; however, it is always nice to have a head start. The results in the following example were built using the "R" programming language, with a publicly-available HMM package whose algorithms are based upon the text "Hidden Markov Models for Time Series," by Walter Zucchini and Iain MacDonald.

## PUTTING HMM TO WORK

One of the great challenges in actuarial work is teasing out the effects of seasonality on claim patterns. In this example, health actuary Albert Franken is having a bit of trouble estimating the number of claims that will be seen in the next month for a small rural clinic. He knows that seasonality effects are significant, as influenza and other maladies spring up in bunches and spike utilization. Fortunately, he has seven years' worth of claim history and a graph of the history readily shows a seasonal pattern in Figure 3:

Upon inspection of the data, Franken theorizes that the clinic's claim activity can be estimated by a Hidden Markov Model with three states:

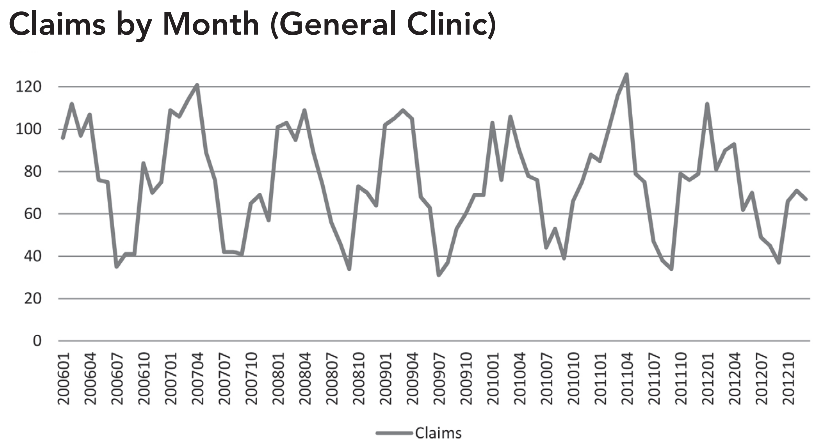
- A low level of claim activity, where claim volume can be modeled as a Poisson distribution with parameter $\lambda_1$.
- A medium level of claim activity, where claim volume can be modeled as a Poisson distribution with parameter $\lambda_2$.
- A high level of claim activity, where claim volume can be modeled as a Poisson distribution with parameter $\lambda_3$.

Estimating the three levels of claim activity at 53 claims (in a low month), 75 claims (in a medium month), and 93 claims (in a high month), Franken uses an EM algorithm to produce a maximum likelihood estimate Hidden Markov Model to describe his clinic's data.

The resulting HMM gives us a transition matrix of:

| | | |
|---|---|---|
| 65.3% | 34.7% | 0.0% |
| 20.3% | 58.1% | 21.5% |
| 0.0% | 30.2% | 69.8% |

**Figure 3**: 7 years of claims data showing a seasonal pattern.




And Poisson parameter values of:

| | |
|---|---|
| Low | 41.74 |
| Med | 71.91 |
| High | 102.89 |

Therefore, the HMM suggests that there are three states of activity—a low level of claims (with about 42 claims per month), a medium level of claims (with about 72 claims per month), and a high level of claims (with about 103 claims per month). Moreover, if a given month is a month with low activity then the following month will either be low as well (with 65 percent probability) or will be medium (with 35 percent probability).

In the last month of clinic data there were 67 medical claims. How can the HMM help our actuary to estimate what could happen next month? He first needs to estimate which activity state we are currently in, and with 67 claims we have a 99.7 percent probability of being in the "medium" level of activity (along with a 0.2 percent probability of being in the "low" level of activity and a 0.1 percent probability of being in the "high" level of activity).

CONTINUED ON **PAGE 20**

Feeding this information into the HMM (after remembering how to do matrix multiplication), Franken finds that next month will be in a "low" level of activity (with 20 percent probability), a "medium" level of activity (with 58 percent probability), or a "high" level of activity (with 22 percent probability). Overall he should expect about 73 claims from the clinic next month:

41.74 * 20% + 71.91 * 58% + 102.89 * 22% = 72.69

## OBSERVING THE UNOBSERVABLE

We hope that you enjoyed our introductory article to Hidden Markov Models. In the December newsletter, we plan on putting together some more in-depth examples in Excel to help inspire your own applications. Stay tuned for more on the most fun models you never saw!

## REFERENCES

Zucchini, Walter and MacDonald, Iain, Hidden Markov Models for Time Series. Upper Chapman & Hill/CRC, 2009. Print. ▼

*Doug Norris*

**Doug Norris,** FSA, MAAA, PhD, is a consulting actuary at Milliman Inc. in Denver, Colo. You can reach him at *doug.norris@milliman.com*

*Brian Grossmiller*

**Brian Grossmiller,** FSA, FCA, MAAA, is provider contract evaluation manager at Moda Health in Portland, Ore. He can be reached at *Brian.Grossmiller@yahoo.com*