Some mathematics behind EFDM

Seija Sirkiä

October 12, 2012

This document briefly describes the application of a Markov model in the current context of forest dynamics, and particularly a suggestion for the estimation of transition probabilities involved.

1 Markov model in rather general terms

A Markov Model considers a set of objects that can exists in a set S of states, called the state space. The objects can individually move from one state to another in discrete time steps. The probability of an object moving from state s_j to state s_i, where $i, j \in \{1, 2, \ldots, M\}$, is given by p_{ij}. Such transistion probabilities can be collected in to an $M \times M$ matrix P which is then called the transition matrix. Note that $\sum_{i=1}^{M} p_{ij} = 1$ must hold.

Given an initial vector of state frequencies $y^{(0)} = (y_1^{(0)}, \ldots, y_M^{(0)})$, the expected frequencies after k time steps is given by $P^k y^{(0)}$.

2 Current context of forest dynamics

In the current context of the EFDM, the objects are unit areas of forest land and the state space consists of combinations of several factors describing such forest areas. These factors include most notably the age and volume classes, which can be visually arranged as an age-volume matrix. The transistions of main interest are transitions within these age-volume matrices as the forests age and gain volume, but the probabilities of these transitions can differ between the levels of other factors involved, such as geographical region, site quality, species composition etc. It is possible to treat this setting as if there was a separate volume transition matrix for each forest type (factor combination) but from a technical point of view the state space involved has an element for each potential combination of age, volume and factor levels. This number, the product of the number of the factor levels (age and volume included), is easily very large.

Managament activities, such as thinning and final felling, introduce another source of different transition probabilities. Here, it is obvious that there should be a different set of transition probabilities per each activity, but again, technically these several conditional probabilities, multiplied by the probability of the activities, sum up to a single transition probability in the end (see section 4).

3 Estimation of transition probabilities

In a broad sense, the best estimator (that uses no outside information) for the probability of an event is the proportion of times that it happens out of total times it is tried and observed, or in other words $s/(s+f)$ where s is the number of successes and f is the number of failures. However, this can be rather unprecise and unreliable when the number of observations is low.

In this particular case, it is very likely that some transition probabilities would have to be estimated with very low observation numbers, or even that there are no observations at all. The number of states gets easily far greater than the number of available sample plots in the available data.

It is possible to retain the intuitive and practical attractiveness of simple proportions and avoid the problem of missing data by using Bayesian estimation in an iterative manner. The idea is to use the information from the data pooled over different factors as a prior and the data from the more finely classified data as observation.

Note that Beta distribution is the conjugate prior for the probability of a Bernoulli trial: this means that the prior distribution is defined by prior parameters α and β and they are updated to $\alpha + s$ and $\beta + f$ after observing s successes and f failures. The Bayes estimate (posterior mean) of the probability is then $(\alpha + s)/(\alpha + \beta + s + f)$.

In the current context, consider first the case of no activity and the transitions from age class i and volume class j, where the volume class is the third largest, so that only three volume transitions – to remain in the current volume class, to move to the next, and to move to two classes up to the highest – are possible. Now, consider one of these three possible volume transitions within a given combination of factor levels. At initial stage use a weak prior $\alpha_0 = 1$, $\beta_0 = 2$. In the data pooled over all factors there are N_1 observations starting from age-volume (i, j) and k_1 of those moved to the volume class under consideration. First order estimate is then given by new parameters $\alpha_1 = 1 + k_1$ and $\beta_1 = 2 + N - k1$ resulting in $(1 + k_1)/(3 + N_1)$. Consider then the division of N_1 among the levels of the most influential factor and pick one of those levels. There are N_2 observations starting from age-volume class (i, j) on that factor level and k_2 of them moved to the volume class considered. Using α_1 and β_1 as prior parameters, we now get $(1+k_1+k_2)/(3+N_1+N_2)$. The process is continued in the same manner for the rest of the factors. Since k_2 is included in k_1 and N_2 is included in N_1, what is obtained in the end is in fact a sum of frequencies from successive divisions, weighted differently: with five factors, the frequency corresponding to a full combination of certain levels of all five factors is counted 5 times, the one corresponding to pooling over the fifth factor is counted 4 times, the one corresponding to pooling over the fifth and fourth 3 times etc. The same process is then repeated for all ages, volumes, factors levels and transitions.

If N_i "runs out", so to speak, at some level this only means that all subdivisions from there on will have the same values of estimate. This goes even for those cases where there are no observations at all, or in other words $N_1 = 0$. These probability estimates will be 1/3. The apparent arbitrary or unintuitive nature of this number should not become an issue because this can only happen in a factor level and age-volume combination that is very rare in practice. Therefore, when the model is used for simulation, these suboptimal estimates are not significantly affecting the end results. In addition, the weak prior can also be chosen such that higher volume gains are less likely, and not equally likely as above.

Instead of a weak prior, it is also possible to use a stronger prior which can be based on for example some growth model or expert knowledge. This is of course the only choice in cases where there are no data available. The advantage of Bayesian estimation described above is the chance to seamlessly combine such knowledge and data.

4 Effect of management activities

The estimation above in principle corresponds to growth in a situation of no management activities. It is straighforwardly applicable also to transition probabilities under any management activity, but when the number of possible transitions goes up, for example when thinning can change not only volume class but, say, dominant species and age class as well, the use of weak priors can produce unacceptable results. It is also likely that the number of managed plots available for the estimation is significantly smaller than the number of unmanaged plots. In those cases some outside information instead of estimation should be used.

In any case, the introduction of activities can be done as follows. For each state s_i (that is, a given combination of volume, age and other factor levels) there is a probability q_{ik} that it receives the activity a_k, with $\sum_{k=0}^{K} q_{ik} = 1$, with K being the number of different actual activites and index 0 referring to no management. The corresponding conditional transition probabilities are given by an $N \times N$ matrix P_k. The complete transition matrix is then $P = \sum_{k=0}^{K} P_k Q_k$, where Q_k is a diagonal matrix with diagonal elements q_{ik}. In other words, the general structure of the model is still of the general Markov form.

In suitable situations it is possible to further divide conditional transition matrix P_k as $P_k = P_0 R_k$. This is interpreted as the activity instantly changing the current state of a unit, after which the unit continues from that state in the same way as an unmanaged unit originally in that state would.

In practice, different scenarios (such as "business as usual" or "schoolbook") are realized through different choices of the set of activities, corresponding activity probabilities Q_k and/or conditional transition probabilities. The structure of the model and consequently of the simulation are not changed.