Logistic Regression on Markov Chains for Crop Rotation Modelling

Matthias C. M. Troffaes
Durham University, UK
matthias.troffaes@gmail.com

Lewis Paton
Durham University, UK
l.w.paton@durham.ac.uk

Abstract

Often, in dynamical systems, such as farmer's crop choices, the dynamics is driven by external non-stationary factors, such as rainfall, temperature, and economy. Such dynamics can be modelled by a non-stationary Markov chain, where the transition probabilities are logistic functions of such external factors. We investigate the problem of estimating the parameters of the logistic model from data, using conjugate analysis with a fairly broad class of priors, to accommodate scarcity of data and lack of strong prior expert opinions. We show how maximum likelihood methods can be used to get bounds on the posterior mode of the parameters.

Keywords. logistic regression, Markov chain, robust Bayesian, conjugate, maximum likelihood, crop

1 Introduction

We wish to accurately model agricultural land use, that is, to predict what crop is grown in any particular field. Usually, farmers follow set patterns of successive yearly crop choices in order to preserve nutrients in the soil. For example, they may have a 3 year cycle, in which they, under normal circumstances, grow wheat for two years, and then leave the field empty for the third year. A very simple model for such crop choices on any particular field is a Markov chain (see for instance [4, 3]), where the state at time i is the crop choice at year i. Such a model makes a simplifying assumption, namely that crop choice in any given year only depends on crop choice in the previous year.

However, crop choices are not only affected by crop choices of the previous year(s): they are also affected by various environmental and economical conditions. In an earlier study, Luo [10] identified some of the most important factors as rainfall, temperature, profit margin, and soil type. To model the effect of these variables on crop choice, in this paper, we propose a logistic regression model for the crop choice transition probabilities. For simplicity, in this paper, we only investigate the impact of rainfall on a simple binary crop choice: wheat, or something else. Generalisation to more than one regressor and to more than two crop choices will be the subject of another paper.

A key challenge with any regression model is to estimate its parameters. First, following [5], we identify a class of conjugate priors for our model. Next, we follow a similar approach to that of the imprecise Dirichlet model [13]: we identify a reasonably vacuous set of conjugate priors, and calculate posterior bounds. A benefit of this approach is that it can also incorporate expert opinion, which will be very useful when studying crop types that are uncommon, such as oats. Our model is thus designed to handle situations in which data is scarce and in which prior expert opinion may be lacking.

The novel contributions of this paper are:

1. We present a first step at including imprecision in non-stationary Markov chains influenced by non-stationary random variables.
2. We propose a novel approach to imprecise logistic regression, based on conjugate analysis.
3. The use of maximum likelihood methods for approximate Bayesian inference in logistic regression, to arrive at fast algorithms when dealing with sets of priors, is new, even though relatively obvious.

The paper is structured as follows. Section 2 introduces the model. Section 3 describes the conjugate prior and posterior distributions, discusses the parameters of the model and their interpretation. Section 4 explains how we can use sets of distributions to obtain posterior bounds. Section 5 has an example. Section 6 concludes the paper, and details future areas of research.

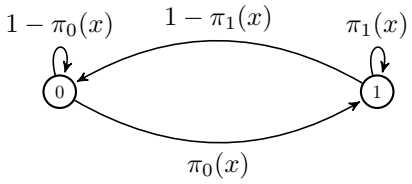

Figure 1: A Markov chain for crop rotations.

2 Logistic Model

We model crop rotations as a non-stationary Markov chain, as depicted in Figure 1.

The model has two states: either our current crop is wheat, which we denote as 1, or our current crop is not wheat, denoted as 0. The transition probabilities in the Markov chain only depend on the previous crop grown and the rainfall—which is where the non-stationarity comes from, as rainfall may change over years. We denote this by:

$$\pi_y(x) \coloneqq P(Y_{i+1} = 1 | Y_i = y, X_i = x) \qquad (1)$$

for all $y \in \{0, 1\}$ and $x \in \mathbb{R}$, where Y_i is the previous crop choice, and X_i is the rainfall recorded just before the planting of crop Y_{i+1}. Note that X_i is not assumed to be part of the state space of the Markov chain, and is simply a non-stationary random variable influencing the transition probabilities.

The impact of rainfall on these transition probabilities is typically either monotonically increasing, or monotonically decreasing. Therefore, a logistic regression model for $\pi_y(x)$ seems fairly reasonable:

$$\pi_y(x) = \frac{e^{\alpha_y + x\beta_y}}{1 + e^{\alpha_y + x\beta_y}}. \qquad (2)$$

where α_y and β_y are parameters of the model.

For example, when it rains a lot, farmers are usually more likely to grow wheat, if the previous crop grown was also wheat; see Figure 2. To produce Figure 2, we used maximum likelihood to fit a logistic regression curve to some actual data when the previous crop grown was wheat—in fact, the data is shown in Table 3 for $y = 1$, and will be explained further in the paper. Note that the relationship is actually reversed if the previous crop is not wheat ($y = 0$).

Also note that the data used here is quite limited, as we used only 10 observations. In reality, wheat versus non-wheat will not be an issue as wheat is a very common crop. However, some crop types, such as oats, are very rare, and will suffer from scarcity of data. For actual applications, our model will be appropriate to handle such crop types specifically. Here, we chose wheat versus non-wheat because that data was readily available, but of course other crop types will be investigated in the future, including rare ones.

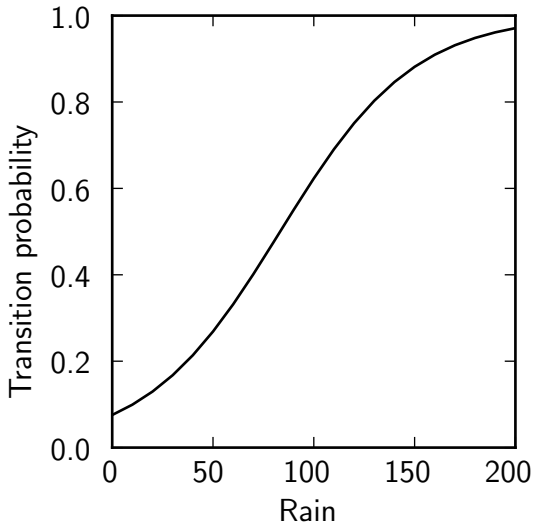

Figure 2: Logistic regression of the probability of growing wheat against rainfall, when the previous crop grown was wheat.

We also assume that we have some model for the regressor X, say a probability density $f_\gamma(x)$ with parameter γ.

For further details about logistic regression, see for instance [1].

3 Parameter Estimation

3.1 Data

We now wish to estimate the parameters of the model, given some data. We have recorded crop transitions and rainfall of a number of fields over a number of years. Specifically, we have $n_y(x)$ observations where the previous crop choice was y and rainfall was x—obviously, $n_y(x)$ will be zero at all but a finite number of $x \in \mathbb{R}$. Of these $n_y(x)$ observations, the crop choice was 1 in $k_y(x)$ cases.

Because we effectively have two separate logistic regression models—one for $y = 0$ and one for $y = 1$—it makes sense to split our data into two sets accordingly. Table 1 tabulates the full data set. Table 2 tabulates the same data, but split according to the value of y.

previous crop y	rain x	current crop total $n_y(x)$	current crop count $k_y(x)$
1	46	1	0
0	52	1	0
0	38	1	1
1	30	1	1
1	37	1	0
⋮	⋮	⋮	⋮

Table 1: Crop rotation data.

previous $y = 0$		
rain x	current crop total $n(x)$	current crop count $k(x)$
52	1	0
38	1	1
⋮	⋮	⋮

previous $y = 1$		
rain x	current crop total $n(x)$	current crop count $k(x)$
46	1	0
30	1	1
37	1	0
⋮	⋮	⋮

Table 2: Crop rotation data split by y.

3.2 Likelihood

Our inspiration is the work by Chen and Ibrahim [5], who propose a conjugate prior distribution of the form:

$$\exp\left(\sum_{i=1}^{m} s\left[t_i(\alpha + x_i\beta) - \ln(1 + e^{\alpha+x_i\beta})\right]\right) \tag{3}$$

where $\vec{x} = (x_1, \ldots, x_m)$ are the observed locations of the regressor, α and β are parameters of the logistic model (as in Eq. (2)), and s and $\vec{t}$ are hyperparameters. However, our notation is simpler if we work directly with the count functions $n_y(x)$ and $k_y(x)$ which are defined for all $x \in \mathbb{R}$, rather than having to enumerate over observed locations explicitly.

Specifically, in terms of $n_y(x)$ and $k_y(x)$, our likelihood is:

$$L_y(\alpha_y, \beta_y, \gamma_y | n_y, k_y) = p_y(k_y | n_y, \alpha_y, \beta_y) f(n_y|\gamma) \tag{4}$$

where

$$f(n_y|\gamma) = \prod_{x\in\mathbb{R}} f_\gamma(x)^{n_y(x)} \tag{5}$$

and

$$p_y(k_y|n_y, \alpha_y, \beta_y) = \prod_{x\in\mathbb{R}} \binom{n_y(x)}{k_y(x)} \pi_y(x)^{k_y(x)} (1 - \pi_y(x))^{n_y(x)-k_y(x)}. \tag{6}$$

The above products over $x \in \mathbb{R}$ are well defined: because $k_y(x)$ and $n_y(x)$ are zero at all but a finite number of x, all but a finite number of factors are equal to one.

Because the likelihood is a product of a function of γ and a function of (α_y, β_y), we can separate our inference procedure accordingly. In the following, we will concern ourselves with inference about (α_y, β_y) only, and leave inference about γ to another paper.

Note that we have subscript y everywhere. To keep notation readable, we will drop it in the remainder of this section. So, we can write:

$$p(k|n, \alpha, \beta) = \prod_{x\in\mathbb{R}} \binom{n(x)}{k(x)} \pi(x)^{k(x)} (1 - \pi(x))^{n(x)-k(x)} \tag{7}$$

For conjugate analysis later, we rewrite this in canonical form [2, p. 202, Definition 4.12], which, after some manipulations, yields:

$$\propto \exp\left(\sum_{x\in\mathbb{R}} k(x)(\alpha + x\beta) - n(x)\ln\left(1 + e^{\alpha+x\beta}\right)\right) \tag{8}$$

up to a normalisation constant that is a function of x only. The above sum over $x \in \mathbb{R}$ is well defined, because $k(x)$ and $n(x)$ are zero at all but a finite number of x.

3.3 Conjugate Prior and Posterior

Following [5, p. 470, Eq. (6.1)], we can now simply define a conjugate prior [2, p. 266, Proposition 5.4] for logistic regression:

$$f_0(\alpha, \beta|s, t) \propto \exp\left(\sum_{x\in\mathbb{R}} s(x)\left[t(x)(\alpha + x\beta) - \ln\left(1 + e^{\alpha+x\beta}\right)\right]\right), \tag{9}$$

where s and t are non-negative functions on $\mathbb{R}$ such that $s(x) = t(x) = 0$ for all but a finite number of $x \in \mathbb{R}$, and $0 \le t(x) \le 1$ for all $x \in \mathbb{R}$.

Writing the posterior distribution down is a simple task [2, p. 269, Proposition 5.5]. We simply multiply

Eq. (8) and Eq. (9), to obtain:

$$
\begin{aligned}
&f(\alpha, \beta | k, n, s, t) \\
&\propto f_0(\alpha, \beta | s, t) p(k, n | \alpha, \beta) && (10) \\
&\propto \exp\left(\sum_{x \in \mathbb{R}} (s(x)t(x) + k(x))(\alpha + x\beta) \right. \\
&\qquad \left. - \, (n(x) + s(x)) \ln\left(1 + e^{\alpha + x\beta}\right) \right) && (11)
\end{aligned}
$$

It is clear the prior distribution and posterior distribution are of the same family:

$$f(\alpha, \beta | k, n, s, t) = f_0(\alpha, \beta | \sigma, \tau) \tag{12}$$

where

$$\sigma(x) \coloneqq s(x) + n(x), \text{ and} \tag{13}$$

$$\tau(x) \coloneqq \frac{s(x)t(x) + k(x)}{s(x) + n(x)}. \tag{14}$$

We now study this family in a bit more detail.

3.4 Interpretation of Hyperparameters

A key problem we are faced with is the choice of prior hyperparameters $s(x)$ and $t(x)$. Ideally we want a direct interpretation of the parameters. Eqs. (13) and (14) show that, as usual, the hyperparameters can be interpreted as a prior virtual sample, with $s(x)$ observations at $X = x$, $s(x)t(x)$ of which are wheat ($Y_{i+1} = 1$). The implications of such specification may however not be entirely clear to an expert, and therefore it seems more appealing, at least to us, to relate the hyperparameters to the prior predictive instead, as is commonly done for the regular exponential family through a famous result by Diaconis and Ylvisaker [6, Theorem 2].

To apply [6, Theorem 2], the number of parameters must be equal to the dimension d of the space $\mathbb{R}^d$ in which the hyperparameter t lives. Therefore, if we relax the model by replacing $\alpha + x\beta$ with an arbitrary function $\theta(x)$—i.e. if we were to drop the assumption that $\pi(x)$ has a logistic form—then [6, Theorem 2] applies, and $t(x)$ is precisely the prior prediction for $\pi(x)$ (see [5, Eqs. (2.4) and (2.5)]).

For our actual model, however, there are only two parameters to estimate (α and β), but unfortunately, the hyperparameter t effectively lives in $\mathbb{R}^d$, where d is the number of x where $s(x)$ is non-zero. Specifically, although we have conjugacy, it is very easy to see that, in general, the prior predictive $\hat{\pi}_0(x)$ is not equal to the hyperparameter $t(x)$, i.e. $t(x)$ is not a prior expectation for $\pi(x)$, unless $d = 2$.

We can still arrive at some sort of interpretation for $t(x)$ as follows. Inspired by [6, Theorem 2], by the usual properties of integration and densities:

$$\iint_{\mathbb{R}^2} \frac{\partial}{\partial \alpha} f_0(\alpha, \beta | s, t) \, \mathrm{d}\alpha \, \mathrm{d}\beta = 0 \tag{15}$$

$$\iint_{\mathbb{R}^2} \frac{\partial}{\partial \beta} f_0(\alpha, \beta | s, t) \, \mathrm{d}\alpha \, \mathrm{d}\beta = 0 \tag{16}$$

These equations yield:

$$\sum_{x \in \mathbb{R}} s(x)t(x) = \sum_{x \in \mathbb{R}} s(x)\hat{\pi}_0(x) \tag{17}$$

$$\sum_{x \in \mathbb{R}} xs(x)t(x) = \sum_{x \in \mathbb{R}} xs(x)\hat{\pi}_0(x) \tag{18}$$

where

$$
\begin{aligned}
\hat{\pi}_0(x) &\coloneqq P(Y_{i+1} = 1 | Y_i = y, X_i = x, s, t) && (19) \\
&= \iint_{\mathbb{R}^2} \pi(x) f_0(\alpha, \beta | s, t) \, \mathrm{d}\alpha \, \mathrm{d}\beta && (20)
\end{aligned}
$$

Note that we should write $\hat{\pi}_{0y}(x)$ but we omit the subscript y for ease of notation as usual.

These equations show that $t(x)$ in some sense 'matches' $\hat{\pi}_0(x)$, the more so for values of x where $s(x)$ is larger. Of course, for any given prior specification of the function $\hat{\pi}_0$, even for fixed s, there will be many different functions t that satisfy Eqs. (17) and (18), so the choice of $t(x)$ is not uniquely determined by our prior expectation about $\pi(x)$.

As mentioned, there is however a special case where the conditions of [6, Theorem 2] are satisfied, and so where we do get a direct interpretation of $t(x)$. This occurs when there are only two points $\{x_1, x_2\}$ where $s(x)$ is non-zero. In this case, Eqs. (17) and (18) do have a unique solution, namely:

$$t(x_1) = \hat{\pi}_0(x_1) \text{ and } t(x_2) = \hat{\pi}_0(x_2) \tag{21}$$

regardless of $s(x_1)$ and $s(x_2)$ (of course, this also follows directly from [6, Theorem 2]). Whence, for simplicity and interpretability, this is the case that we will consider in practical examples later. In this case, as we shall see, $s(x_1)$ and $s(x_2)$ also carry their usual interpretation, in determining the speed by which our posterior will move away from our prior.

4 Inference

4.1 Posterior Transition Probability

For inference, we are mostly interested in the posterior transition probability:

$$
\begin{aligned}
\hat{\pi}(x) &\coloneqq P(Y_{i+1} = 1 | Y_i = y, X_i = x, k, n, s, t) && (22) \\
&= \iint_{\mathbb{R}^2} \pi(x) f(\alpha, \beta | k, n, s, t) \, \mathrm{d}\alpha \, \mathrm{d}\beta && (23)
\end{aligned}
$$

where it is worth recalling that $\pi(x)$ is a non-linear (logistic) function of α and β. Specifically, taking into account our uncertainty about α and β as given by the posterior, we are interested in evaluating Eq. (23). The challenge now is the evaluation of the integral. One option is to directly numerically integrate. However, as eventually, we want to use sets of distributions, this may not necessarily be the most sensible route to take.

Therefore, we may prefer to rely on faster approximations of the integral. A first crude idea would be to approximate the prior (Eq. (9)) by a multivariate normal distribution; Chen and Ibrahim [5] mention that for large sample sizes this approximation yields the exact solution. Whilst the mean can be easily approximated through the mode, obtaining the covariance structure is somewhat more difficult (a starting point would be [5, Theorem 2.3]). Interestingly, there are variational techniques for direct updating of the mean and covariance structure [9], which means that we would need to perform the multivariate normal approximation only once, on the initial prior.

However, this approach still requires numerical integration. As just mentioned, when we move to sets of priors, this might easily become computationally intractable, as we will have to update, approximate, and integrate, for every prior in the set. A more crude but also much faster approximation would be to simply pretend that all probability mass is concentrated at the mode of the posterior. It is relatively straightforward to show that the mode can be obtained by solving the following system of non-linear equations for α and β:

$$\sum_{x\in\mathbb{R}} \sigma(x)\tau(x) = \sum_{x\in\mathbb{R}} \sigma(x)\pi(x) \tag{24}$$

$$\sum_{x\in\mathbb{R}} x\sigma(x)\tau(x) = \sum_{x\in\mathbb{R}} x\sigma(x)\pi(x) \tag{25}$$

where it is again worth recalling that $\pi(x)$ is a non-linear (logistic) function of α and β. To obtain an approximate value for $\hat{\pi}(x)$, we simply plug in the solution (α^*, β^*) into the expression for $\pi(x)$ (see Eq. (2)):

$$\hat{\pi}(x) \approx \frac{e^{\alpha^*+x\beta^*}}{1+e^{\alpha^*+x\beta^*}}. \tag{26}$$

Although this approximation is obviously horribly crude, we note that in fact it corresponds to the maximum likelihood estimate, where the data has been augmented with pseudo counts. Hence, it reflects current practice quite well, and arguably even improves it, by allowing for additional prior information to be taken into account.

Solving a system of non-linear equations is non-trivial. However, Green [8] provides a Newton Raphson algorithm specifically for the maximum likelihood estimate of logistic regression. We can essentially recycle algorithms like these to find the mode, simply by adding some pseudo counts to the data to reflect our prior.

4.2 Sets of Prior Distributions

We now want to propose sets of prior distributions, in a similar vein to Walley's imprecise Dirichlet Model [13]. In this section, we study the inferences resulting from an arbitrary but fixed prior function for $s(x)$, namely:

$$s(x) \coloneqq \begin{cases} s & \text{if } x \in \mathcal{X}, \\ 0 & \text{otherwise,} \end{cases} \tag{27}$$

for some finite set $\mathcal{X} \subseteq \mathbb{R}$, and an arbitrary set of prior functions $\mathfrak{T}$ for $t(x)$. We explain how to calculate posterior bounds based on this set of priors, and the observed data. Practical choices for reasonably vacuous sets of prior distributions will be discussed further in Section 5.

4.3 Posterior Transition Probability Bounds

For the above choice of $s(x)$, Eqs. (24) and (25) can be written as:

$$s\sum_{x\in\mathcal{X}}(\pi(x)-t(x)) + \sum_{x\in\mathbb{R}}(n(x)\pi(x)-k(x)) = 0, \tag{28}$$

$$s\sum_{x\in\mathcal{X}} x(\pi(x)-t(x)) + \sum_{x\in\mathbb{R}} x(n(x)\pi(x)-k(x)) = 0. \tag{29}$$

If we can solve the above equations for all $t \in \mathfrak{T}$, then we obtain a set Θ^* of solutions (α^*, β^*), one solution for each $t \in \mathfrak{T}$. Each member of Θ^* corresponds to an estimate of the posterior transition probability as in Eq. (26). Whence,

$$\underline{\pi}(x) \approx \inf_{(\alpha^*,\beta^*)\in\Theta^*} \frac{e^{\alpha^*+x\beta^*}}{1+e^{\alpha^*+x\beta^*}}, \tag{30}$$

$$\overline{\pi}(x) \approx \sup_{(\alpha^*,\beta^*)\in\Theta^*} \frac{e^{\alpha^*+x\beta^*}}{1+e^{\alpha^*+x\beta^*}}, \tag{31}$$

are the desired lower and upper posterior approximations of the transition probability.

5 Example

As discussed in Section 3.4, there is a direct interpretation of $t(x)$ when $\mathcal{X} = \{x_1, x_2\}$. We will explore this case here.

previous $y = 0$		
rain x	current crop total $n(x)$	current crop count $k(x)$
18	1	1
68	1	1
24	1	1
19	1	1
99	1	0
16	1	0
20	1	0
119	1	0
102	1	0
87	1	1
17	1	0
29	1	0

previous $y = 1$		
rain x	current crop total $n(x)$	current crop count $k(x)$
72	1	1
105	1	1
6	1	0
104	1	1
77	1	0
69	1	0
15	1	0
63	1	0
35	1	1
25	1	0

Table 3: Actual crop rotation data split by y.

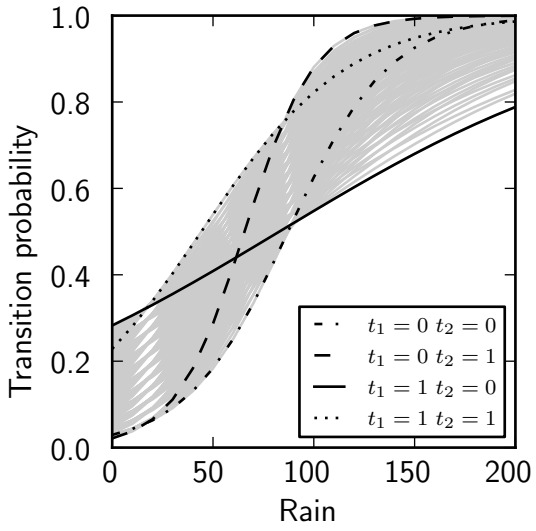

Figure 3: $\hat{\pi}(x)$ based on $\mathfrak{T}_v$ and $\mathfrak{T}'_v$ for $y = 1$.

We take a set of functions for $t(x)$ and a constant s. The most vacuous choice would be:

$$\mathfrak{T}_v = \{t \in \mathbb{R}^{\mathbb{R}} \colon t(x) = 0 \text{ when } x \notin \mathcal{X}, \\ 0 < t(x) < 1 \text{ when } x \in \mathcal{X}\} \quad (32)$$

Solving the optimisation problem (Eqs. (30) and (31)) over $\mathfrak{T}_v$ is rather involved. For a simple quick analysis, we restrict ourselves to the extreme points of $\mathfrak{T}_v$, namely:

$$\mathfrak{T}'_v = \{t \in \mathbb{R}^{\mathbb{R}} \colon t(x) = 0 \text{ when } x \notin \mathcal{X}, \\ (t(x_1), t(x_2)) \in \{(0,0),(0,1),(1,0),(1,1)\}\} \quad (33)$$

We will use data collected from actual fields to illustrate the ideas we have talked about. The data is shown in Table 3. It consists of 22 observations of crop transitions [12], and the corresponding rainfall recorded in the month of planting for each crop [11].

Figure 3 shows $\hat{\pi}(x)$ for each element of $\mathfrak{T}'_v$, where we have specified $s = 2$, $x_1 = 30$, and $x_2 = 80$, and we are looking at the model for $y = 1$. Each line corresponds to one estimate. The grey region represents the posterior estimates from the most vacuous set $\mathfrak{T}_v$ (we actually used a 21×21 grid over the unit square). As can be seen $\underline{\pi}$ and $\overline{\pi}$ are very closely matched in both cases (almost shockingly so!), so it seems very reasonable to use only $\mathfrak{T}'_v$ instead of the full set $\mathfrak{T}_v$, for ease of computation.

Note that the case $t_1 = 1$, $t_2 = 0$, goes against the data, and corresponds to a non-natural shape for π in this problem. Thus, Figure 3 also highlights the importance of including constraints on π which follow from prior expert opinion, for instance by removing those values from $\mathfrak{T}_v$ for which π violates those constraints. A less vacuous prediction for larger values of x would result. Of course, other techniques for learning under order constraints, which have been studied for instance in the context of Bayesian network learning [7], could also have some potential here.

Our choice for x_1 and x_2 is also important. In Figure. 4, we use $x_1 = 10$ and $x_2 = 100$, which lean more towards the extremes of the range of observations in x. This changes our inferences quite substantially. The largest impact is observed for the case $t_1 = 1$, $t_2 = 0$, and as we just saw, removing such unnatural values for t from $\mathfrak{T}_v$ might be reasonable. In any case, this also demonstrates the importance of choosing x_1 and x_2 sensibly, particularly under the vacuous model $\mathfrak{T}_v$. For example, a sensible choice would be to take for x_1 the first quartile and for x_2 the third quartile, of the observations in x (or of our prior distribution for x).

The inference also depends on the the value of s. As in the imprecise Dirichlet model, smaller values of s produce tighter bounds, as seen in Figure. 5.

6 Conclusion

In this paper, we proposed a new model for land use, which aims to properly capture epistemic uncertainty about crop rotations, in an interpretable, robust, and efficient way. Thereby, we presented a first step at including imprecision in non-stationary Markov chains influenced by non-stationary random variables.

In a nutshell, starting from earlier work by Chen and

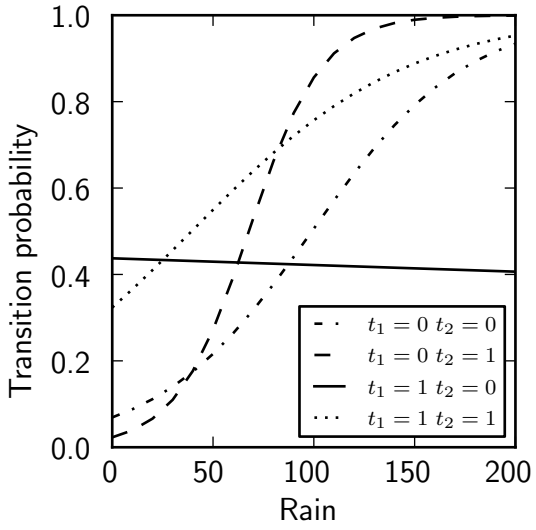

Figure 4: The impact of changing x_1 and x_2.

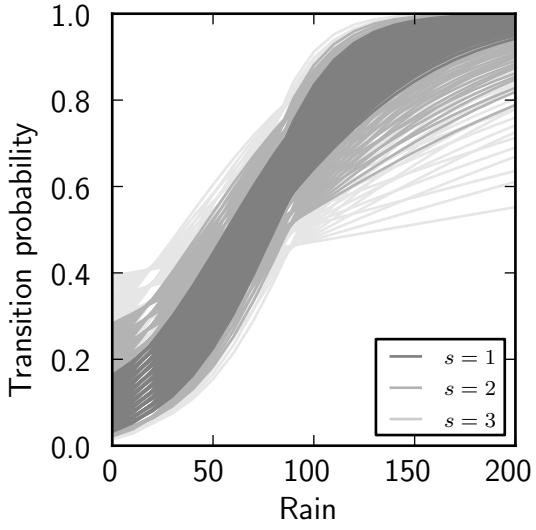

Figure 5: 3 different sets of $\hat{\pi}(x)$ for different values of s.

Ibrahim [5], we proposed a new model for imprecise logistic regression, using sets of conjugate prior distributions for a generalised linear model with logistic link function, to get bounds on the probability of growing wheat as a function of rainfall.

We investigated the interpretation of the hyperparameters of the model, which turns out to be somewhat non-trivial, unless the model is constrained in a very specific way.

We care about robustness, because typically, for certain rare crop types, only a small amount of data is available. This results in posterior probabilities which are highly sensitive to prior specifications. By using sets of priors, our approach allows us to draw accurate robust inferences even from near-vacuous prior knowledge about crop rotations.

Due to the non-linearity of our model, one might fear that the updating process is highly complicated. We proposed the use of maximum likelihood methods for approximate Bayesian inference, effectively via data augmentation, to arrive at fast algorithms when dealing with sets of priors. Much to our surprise, it turns out that using the set of extreme points in our prior specifications still captures the posterior bounds extremely well. We suspect that this is due to the monotonicity of the link function.

An obvious weakness of our analysis is the use of the posterior mode as a very crude approximation to the actual posterior expectation. However, the other options for evaluating the posterior expectation are computationally far more complex, making a robust analysis over sets of parameter values infeasible, at least in our initial attempts. Nevertheless, the use of the posterior mode corresponds quite well to current practice: a standard technique for estimating the parameters in logistic regression goes by maximum likelihood estimation, and the posterior mode can be interpreted as such.

Concerning the actual crop modelling, this work is still in its infancy. We have yet to judge the effects of the simplifying assumptions we have made, and we still need to assess the validity of our model. We plan to use the posterior bounds in conjunction with a predictive model for rainfall, to make predictions about future crop distributions. We also plan to extend the model to deal with multiple crop choices (i.e. more than just wheat) and multiple regressors (i.e. not just rainfall, but also economic factors).

Acknowledgements

The second author is supported by the Food and Environmental Research Agency (FERA), and EPSRC.

We thank Andy Hart and Nigel Boatman from FERA for technical advice. Finally, we are indebted to all three reviewers for their comments which have improved the paper considerably.

References

[1] Alan Agresti. *Categorical Data Analysis*. John Wiley and Sons, second edition, 2002.

[2] Jose M. Bernado and Adrian F. M. Smith. *Bayesian Theory*. John Wiley and Sons, 1994.

[3] M. S. Castellazzi, J. Matthews, F. Angevin, C. Sausse, G. A. Wood, P. J. Burgess, I. Brown, K. F. Conrad, and J. N. Perry. Simulation scenarios of spatio-temporal arrangement of crops at the landscape scale. *Environmental Modelling and Software*, 25(12):1881–1889, 2010.

[4] M. S. Castellazzi, G. A. Wood, P. J. Burgess, J. Morris, K. F. Conrad, and J. N. Perry. A systematic representation of crop rotations. *Agricultural Systems*, 97:26–33, 2008.

[5] Ming-Hui Chen and Joseph G. Ibrahim. Conjugate priors for generalized linear models. *Statistica Sinica*, 13:461–476, 2003.

[6] Persi Diaconas and Donald Ylvisaker. Conjugate priors for exponential families. *The Annals of Statistics*, 7(2):269–281, 1979.

[7] Ad Feelders and Linda C. van der Gaag. Learning Bayesian network parameters under order constraints. *International Journal of Approximate Reasoning*, 42:37–53, May 2006.

[8] P. J. Green. Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society Series B*, 46(2):149–192, 1984.

[9] Tommi S. Jaakkola and Michael I. Jordan. A variational approach to Bayesian logistic regression models and their extensions. In *Sixth International Workshop on Artificial Intelligence and Statistics*, 1997.

[10] Weiqi Luo. Land use modelling. Internal Report, Food and Envirnoment Research Agency, 2010.

[11] Data collected by the Met Office. `http://www.metoffice.gov.uk/climate/uk/stationdata/`. Accessed: 11/02/2013.

[12] Data collected by the Rural Payments Agency under the integrated administration and control system for the administration of subsidies under the common agricultural policy.

[13] Peter Walley. Inferences from multinomial data: Learning about a bag of marbles. *Journal of the Royal Statistical Society, Series B*, 58(1):3–34, 1996.