

# A model for animal behaviour which incorporates feedback

---

## 16.1 Introduction

Animal behaviourists are interested in the causal factors that determine behavioural sequences — i.e. when animals perform particular activities, and under which circumstances they switch to alternative activities. It is accepted that observed behaviour results from the nervous system integrating information regarding the physiological state of the animal, e.g. the levels of nutrients in the blood, with sensory inputs, e.g. concerning the levels of nutrients in a food (Barton Browne, 1993). The combined physiological and perceptual state of the animal is termed the ‘motivational state’ (McFarland, 1999). MacDonald and Raubenheimer (1995) modelled behaviour sequences using an HMM whose unobserved underlying states were interpreted as motivational states. Their model captures an important aspect of the causal structure of behaviour, since an animal in a given motivational state (e.g. hungry) might perform not only the most likely behaviour for that state (feed) but also other behaviours (groom, drink, walk, etc.). There is not a one-to-one correspondence between motivational state and behaviour. And it is the runlength distributions of the motivational states that are of interest, rather than those of the observed behaviours. HMMs do not, however, take into account the fact that, in many cases, behaviour also influences motivational state; feeding, for example, leads to satiation.

We describe here a model, proposed by Zucchini, Raubenheimer and MacDonald (2008), that incorporates such feedbacks, and we apply it in order to model observed feeding patterns of caterpillars. We define a ‘nutrient level’, which is determined by the animal’s recent feeding behaviour and which, in turn, influences the probability of transition to a different motivational state. Latent-state models, including HMMs, provide a means of grouping two or more behaviours, such as feeding and grooming, into ‘activities’. Here the activities of interest are meal-taking, an activity characterized by feeding interspersed by brief pauses, and inter-meal intervals, in which the animal mainly rests but might also feed for brief periods.

The proposed model is not an HMM because the states do not form

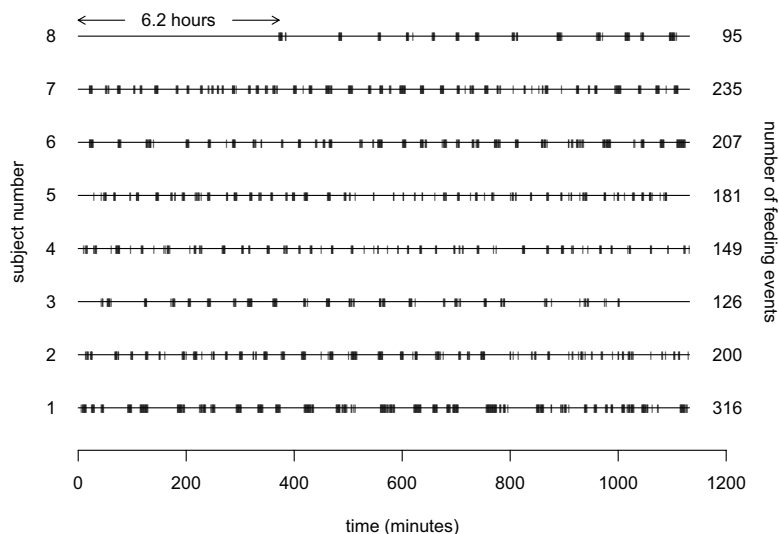


Figure 16.1 *Feeding behaviour of eight Helicoverpa armigera caterpillars observed at 1-minute intervals.*

a Markov chain. It can be regarded as a special-purpose extension of HMMs, and shares many of the features of HMMs. We therefore discuss the theoretical aspects of the model before moving on to an application.

In Section 16.9 we demonstrate the application of the model to data collected in an experiment in which eight caterpillars were observed at one-minute intervals for almost 19 hours, and classified as feeding or not feeding. The data are displayed in Figure 16.1.

## 16.2 The model

Suppose an animal is observed at integer times  $t = 1, 2, \dots, T$ , and classified as feeding at time  $t$  ( $X_t = 1$ ) or not ( $X_t = 0$ ). We propose the following model.

There are two possible (unobserved) motivational states, provisionally labelled ‘hungry’ (state 1) and ‘sated’ (state 2). The state process  $\{C_t\}$  is a process in which the transition probabilities are driven by a process  $\{N_t\}$  termed the ‘nutrient level’, which takes values in  $[0, 1]$ . We assume that  $N_t$  is some function of  $N_{t-1}$  and  $X_t$ . The development that follows is applicable more generally, but we shall restrict our attention to the

exponential filter:

$$N_t = \lambda X_t + (1 - \lambda)N_{t-1} \quad (t = 1, 2, \dots, T). \quad (16.1)$$

In biological terms,  $N_t$  would correspond approximately with the levels of nutrients in the blood (Simpson and Raubenheimer, 1993). We take the transition probabilities

$$\gamma_{ij}(n_t) = \Pr(C_{t+1} = j \mid C_t = i, N_t = n_t), \quad (t = 1, 2, \dots, T-1)$$

to be determined as follows by the nutrient level:

$$\text{logit } \gamma_{11}(n_t) = \alpha_0 + \alpha_1 n_t, \quad \text{logit } \gamma_{22}(n_t) = \beta_0 + \beta_1 n_t; \quad (16.2)$$

and by the row-sum constraints:

$$\gamma_{11}(n_t) + \gamma_{12}(n_t) = 1, \quad \gamma_{21}(n_t) + \gamma_{22}(n_t) = 1.$$

In Equation (16.2) the logit could be replaced by some other monotonic function  $g : (0, 1) \rightarrow \mathbb{R}$ .

In state 1, the probability of feeding is always  $\pi_1$ , regardless of earlier motivational state, behaviour or nutrient level; similarly  $\pi_2$  in state 2. The behaviour  $X_t$  influences the nutrient level  $N_t$ , which in turn determines the transition probabilities of the state process and so influences the state occupied at the next time point,  $t + 1$ .

Our two fundamental assumptions are as follows, for  $t = 2, 3, \dots, T$  and  $t = 1, 2, \dots, T$  respectively:

$$\Pr(C_t \mid \mathbf{C}^{(t-1)}, N_0, \mathbf{X}^{(t-1)}) = \Pr(C_t \mid C_{t-1}, N_{t-1}); \quad (16.3)$$

and

$$\begin{aligned} \Pr(X_t = 1 \mid \mathbf{C}^{(t)}, N_0, \mathbf{X}^{(t-1)}) &= \Pr(X_t = 1 \mid C_t) \\ &= \begin{cases} \pi_1 & \text{if } C_t = 1 \\ \pi_2 & \text{if } C_t = 2. \end{cases} \end{aligned} \quad (16.4)$$

We use the convention that  $\mathbf{X}^{(0)}$  is an empty set of random variables, and similarly for  $\mathbf{N}^{(0)}$ .

Since — given the parameter  $\lambda$  —  $\mathbf{N}^{(t-1)}$  is completely determined by  $N_0$  and  $\mathbf{X}^{(t-1)}$ , the above two assumptions can equivalently be written as

$$\Pr(C_t \mid \mathbf{C}^{(t-1)}, N_0, \mathbf{N}^{(t-1)}, \mathbf{X}^{(t-1)}) = \Pr(C_t \mid C_{t-1}, N_{t-1});$$

and

$$\Pr(X_t = 1 \mid \mathbf{C}^{(t)}, N_0, \mathbf{N}^{(t-1)}, \mathbf{X}^{(t-1)}) = \begin{cases} \pi_1 & \text{if } C_t = 1 \\ \pi_2 & \text{if } C_t = 2. \end{cases}$$

We expect  $\pi_1$  to be close to 1 and  $\pi_2$  to be close to 0. If we define

$$p_i(x) = \Pr(X_t = x \mid C_t = i) = \pi_i^x (1 - \pi_i)^{1-x} \quad (\text{for } x = 0, 1; i = 1, 2),$$

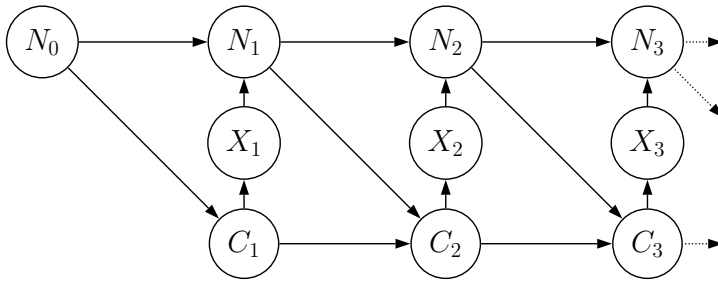


Figure 16.2 *Directed graph representing animal behaviour model.*

assumption (16.4) becomes

$$\Pr(X_t = x \mid \mathbf{C}^{(t)}, N_0, \mathbf{X}^{(t-1)}) = p_{c_t}(x).$$

The model is conveniently represented by the directed graph in Figure 16.2. Notice that there is a path from  $C_1$  to  $C_3$  that does not pass through  $C_2$ ; the state process is therefore not in general a Markov process.

The following are treated as parameters of the model:  $\alpha_0, \alpha_1, \beta_0, \beta_1, \pi_1, \pi_2$ , and  $\lambda$ . In addition, we take  $n_0 \in [0, 1]$  to be a parameter, albeit not one of any intrinsic interest.

One can also treat as a parameter the distribution of  $C_1$  given  $N_0$  (which we denote by the vector  $\boldsymbol{\delta}$ ), or more precisely, the probability  $\delta_1 = \Pr(C_1 = 1 \mid N_0)$ . It can, however, be shown that on maximizing the resulting likelihood one will simply have either  $\hat{\delta}_1 = 1$  or  $\hat{\delta}_1 = 0$ . We follow this approach, but in the application described in Section 16.9 it turns out in any case that it is biologically reasonable to assume that  $\delta_1 = 0$ , i.e. that the subject always starts in state 2.

### 16.3 Likelihood evaluation

Given a sequence of observations  $\{\mathbf{x}^{(T)}\}$  assumed to arise from such a model, and given values of the parameters listed above, we need to be able to evaluate the likelihood, both to estimate parameters and to find the marginal and conditional distributions we shall use in this analysis. First we write the likelihood as a  $T$ -fold multiple sum, then we show that this sum can be efficiently computed via a recursion.

### 16.3.1 The likelihood as a multiple sum

We therefore seek

$$L_T = \Pr(\mathbf{X}^{(T)} \mid N_0) = \sum_{c_1, \dots, c_T} \Pr(\mathbf{C}^{(T)}, \mathbf{X}^{(T)} \mid N_0).$$

The summand  $\Pr(\mathbf{C}^{(T)}, \mathbf{X}^{(T)} \mid N_0)$  may be decomposed as follows:

$$\begin{aligned} \Pr(\mathbf{C}^{(T)}, \mathbf{X}^{(T)} \mid N_0) &= \Pr(C_1 \mid N_0) \Pr(X_1 \mid C_1, N_0) \times \\ &\quad \prod_{t=2}^T \left( \Pr(C_t \mid \mathbf{C}^{(t-1)}, N_0, \mathbf{X}^{(t-1)}) \Pr(X_t \mid \mathbf{C}^{(t)}, N_0, \mathbf{X}^{(t-1)}) \right) \\ &= \Pr(C_1 \mid N_0) \Pr(X_1 \mid C_1) \prod_{t=2}^T \left( \Pr(C_t \mid C_{t-1}, N_{t-1}) \Pr(X_t \mid C_t) \right). \end{aligned}$$

The first equality follows by repeated application of the definition of conditional probability, and the second from assumptions (16.3) and (16.4). We therefore conclude that

$$L_T = \sum_{c_1, \dots, c_T} \left( \delta_{c_1} p_{c_1}(x_1) \prod_{t=2}^T \left( \gamma_{c_{t-1}, c_t}(n_{t-1}) p_{c_t}(x_t) \right) \right).$$

### 16.3.2 Recursive evaluation

The likelihood is therefore a sum of the form

$$L = \sum_{c_1=1}^2 \sum_{c_2=1}^2 \dots \sum_{c_T=1}^2 \left( \alpha_1(c_1) \prod_{t=2}^T f_t(c_{t-1}, c_t) \right); \quad (16.5)$$

that is, if here we define

$$\alpha_1(j) = \delta_j p_j(x_1)$$

and, for  $t = 2, 3, \dots, T$ ,

$$f_t(i, j) = \gamma_{ij}(n_{t-1}) p_j(x_t).$$

Multiple sums of the form (16.5) can in general be evaluated recursively; see Exercise 7 of Chapter 2. Indeed, this is the key property of hidden Markov likelihoods that makes them computationally feasible, and here makes it unnecessary to sum explicitly over the  $2^T$  terms. (See Lange (2002, p. 120) for a very useful discussion of such recursive summation as applied in the computation of likelihoods of pedigrees, a more complex problem than the one we need to consider.)

From Exercise 7 of Chapter 2 (part (b)) we conclude that

$$L = \boldsymbol{\alpha}_1 \mathbf{F}_2 \mathbf{F}_3 \cdots \mathbf{F}_T \mathbf{1}',$$

where the  $2 \times 2$  matrix  $\mathbf{F}_t$  has  $(i, j)$  element equal to  $f_t(i, j)$ , and  $\boldsymbol{\alpha}_1$  is the vector with  $j$ th element  $\alpha_1(j)$ . In the present context,  $\boldsymbol{\alpha}_1 = \boldsymbol{\delta} \mathbf{P}(x_1)$ ,  $\mathbf{F}_2 = \boldsymbol{\Gamma}(n_1) \mathbf{P}(x_2)$ , and similarly  $\mathbf{F}_3$ , etc. Hence the likelihood of the model under discussion can be written as the matrix product

$$L_T = \boldsymbol{\delta} P(x_1) \boldsymbol{\Gamma}(n_1) \mathbf{P}(x_2) \boldsymbol{\Gamma}(n_2) \cdots \boldsymbol{\Gamma}(n_{T-1}) \mathbf{P}(x_T) \mathbf{1}', \quad (16.6)$$

where the vector  $\boldsymbol{\delta}$  is the distribution of  $C_1$  given  $N_0$ ,  $\mathbf{P}(x_t)$  is the diagonal matrix with  $i$ th diagonal element  $p_i(x_t)$ , and  $\boldsymbol{\Gamma}(n_t)$  is the matrix with  $(i, j)$  element  $\gamma_{ij}(n_t)$ .

As usual, precautions have to be taken against numerical underflow, but otherwise the matrix product (16.6) can be used as it stands to evaluate the likelihood. The computational effort is linear in  $T$ , the length of the series, in spite of the fact that there are  $2^T$  terms in the sum.

## 16.4 Parameter estimation by maximum likelihood

Estimation may be carried out by direct numerical maximization of the log-likelihood. Since the parameters  $\pi_1$ ,  $\pi_2$ ,  $\lambda$ ,  $n_0$  and  $\delta_1$  are constrained to lie between 0 and 1, it is convenient to reparametrize the model in order to avoid these constraints. A variety of methods were used in this work to carry out the optimization and checking: the Nelder–Mead simplex algorithm, simulated annealing and methods of Newton type, as implemented by the **R** functions `optim` and `nlm`.

An alternative to direct numerical maximization would be to use the EM algorithm. In this particular model, however, there are no closed-form expressions for the parameter estimates given the complete data, i.e. given the observations plus the states occupied at all times. The M step of the EM algorithm would therefore involve numerical optimization, and it seems circuitous to apply an algorithm which requires numerical optimization in each iteration, instead of only once. Whatever method is used, however, one has to bear in mind that there may well be multiple local optima in the likelihood.

## 16.5 Model checking

When we have fitted a model to the observed behaviour of an animal, we need to examine the model to assess its suitability. One way of doing so is as follows.

It is a routine calculation to find the forecast distributions under the fitted model, i.e. the distribution of each  $X_t$  given the history  $\mathbf{X}^{(t-1)}$ .

The probabilities are the ratios of two likelihood values:

$$\Pr(X_t | \mathbf{X}^{(t-1)}) = \Pr(\mathbf{X}^{(t)}) / \Pr(\mathbf{X}^{(t-1)}) = L_t / L_{t-1}.$$

We denote by  $\hat{p}_t$  the probability  $\Pr(X_t = 1 | \mathbf{X}^{(t-1)})$  computed thus under the model. Since the joint probability function of  $\mathbf{X}^{(T)}$  factorizes as follows:

$$\Pr(\mathbf{X}^{(T)}) = \Pr(X_1) \Pr(X_2 | X_1) \Pr(X_3 | \mathbf{X}^{(2)}) \cdots \Pr(X_T | \mathbf{X}^{(T-1)}),$$

we have a problem of the following form. There are  $T$  binary observations  $x_t$ , assumed to be drawn independently, with  $E(x_t) = \hat{p}_t$ . We wish to test the null hypothesis  $E(x_t) = \hat{p}_t$  (for all  $t$ ), or equivalently

$$H_0 : g(E(x_t)) = g(\hat{p}_t),$$

where  $g$  is the logit transform. We consider an alternative hypothesis of the form

$$H_A : g(E(x_t)) = f(g(\hat{p}_t)),$$

where  $f$  is a smoothing spline (see e.g. Hastie and Tibshirani (1990)). Departure of  $f$  from the identity function constitutes evidence against the null hypothesis, and a plot of  $f$  against the identity function will reveal the nature of the departure.

## 16.6 Inferring the underlying state

A question that is of interest in many applications of latent-state models is this: what are the states of the latent process (here  $\{C_t\}$ ) that are most likely (under the fitted model) to have given rise to the observation sequence? This is the decoding problem discussed in Section 5.3. More specifically, ‘local decoding’ of the state at time  $t$  refers to the determination of the state  $i_t$  which is most likely at that time, i.e.

$$i_t = \operatorname{argmax}_{i=1,2} \Pr(C_t = i | \mathbf{X}^{(T)} = \mathbf{x}^{(T)}).$$

In the context of feeding, local decoding might be of interest for determining the specific sets of sensory and metabolic events that distinguish meal-taking from intermeal breaks (Simpson and Raubenheimer, 1993).

In contrast, global decoding refers to the determination of that sequence of states  $c_1, c_2, \dots, c_T$  which maximizes the conditional probability

$$\Pr(\mathbf{C}^{(T)} = \mathbf{c}^{(T)} | \mathbf{X}^{(T)} = \mathbf{x}^{(T)});$$

or equivalently the joint probability

$$\Pr(\mathbf{C}^{(T)}, \mathbf{X}^{(T)}) = \delta_{c_1} \prod_{t=2}^T \gamma_{c_{t-1}, c_t}(n_{t-1}) \prod_{t=1}^T p_{c_t}(x_t).$$

Global decoding can be carried out, both here and in other contexts, by means of the Viterbi algorithm: see Section 5.3.2. For the sake of completeness we present the details here, although little is needed beyond that which is in Section 5.3.2. Define

$$\xi_{1i} = \Pr(C_1 = i, X_1 = x_1) = \delta_i p_i(x_1),$$

and, for  $t = 2, 3, \dots, T$ ,

$$\xi_{ti} = \max_{c_1, c_2, \dots, c_{t-1}} \Pr(\mathbf{C}^{(t-1)} = \mathbf{c}^{(t-1)}, C_t = i, \mathbf{X}^{(T)} = \mathbf{x}^{(T)}).$$

It can then be shown that the probabilities  $\xi_{tj}$  satisfy the following recursion, for  $t = 2, 3, \dots, T$ :

$$\xi_{tj} = \left( \max_{i=1,2} \left( \xi_{t-1,i} \gamma_{ij}(n_{t-1}) \right) \right) p_j(x_t). \quad (16.7)$$

This provides an efficient means of computing the  $T \times 2$  matrix of values  $\xi_{tj}$ , as the computational effort is linear in  $T$ . The required sequence of states  $i_1, i_2, \dots, i_T$  can then be determined recursively from

$$i_T = \operatorname{argmax}_{i=1,2} \xi_{Ti}$$

and, for  $t = T - 1, T - 2, \dots, 1$ , from

$$i_t = \operatorname{argmax}_{i=1,2} \left( \xi_{ti} \gamma_{i,i_{t+1}}(n_t) \right).$$

## 16.7 Models for a heterogeneous group of subjects

There are several directions in which the model may be extended if that is useful for the application intended. For instance, we may wish to investigate the effects of subject- or time-specific covariates on feeding behaviour, in which case we would need to model the subjects as a group rather than as individuals.

### 16.7.1 Models assuming some parameters to be constant across subjects

One, fairly extreme, model for a group of  $I$  subjects would be to assume that they are independent and that, apart from nuisance parameters, they have the same set of parameters, i.e. the seven parameters  $\alpha_0, \alpha_1, \beta_0, \beta_1, \pi_1, \pi_2$  and  $\lambda$  are common to the  $I$  subjects. The likelihood is in this case just the product of the  $I$  individual likelihoods, and is a function of the seven parameters listed,  $I$  values of  $n_0$  and (and if they are treated as parameters)  $I$  values of  $\delta_1$ ;  $7 + 2I$  parameters in all. At the other extreme is the collection of  $I$  individual models for the subjects, each of which has its own set of nine parameters and is assumed independent of the other subjects. In this case the comparable number



of parameters is  $9I$ . Intermediate between these two cases are models that assume that some but not all of the seven parameters listed are common to the  $I$  subjects. For instance, one might wish to assume that only the probabilities  $\pi_1$  and  $\pi_2$  are common to all subjects.

### 16.7.2 Mixed models

However, in drawing overall conclusions from a group, it may be useful to allow for between-subject variability by some means other than merely permitting parameters to differ between subjects. One way of doing so is to incorporate random effects; another is to use subject-specific covariates.

The incorporation of a single subject-specific random effect into a model for a group of subjects is in principle straightforward; see Altman (2007) for a general discussion of the introduction of random effects into HMMs. For concreteness, suppose that the six parameters  $\alpha_0$ ,  $\alpha_1$ ,  $\beta_0$ ,  $\beta_1$ ,  $\pi_1$  and  $\pi_2$  can reasonably be supposed constant across subjects, but not  $\lambda$ . Instead we suppose that  $\lambda$  is generated by some density  $f$  with support  $[0,1]$ .

Conditional on  $\lambda$ , the likelihood of the observations on subject  $i$  is

$$L_T(i, \lambda) = \delta \mathbf{P}(x_{i1}) \Gamma(n_{i1}) \mathbf{P}(x_{i2}) \Gamma(n_{i2}) \cdots \Gamma(n_{iT-1}) \mathbf{P}(x_{iT}) \mathbf{1}',$$

where  $\{x_{it} : t = 1, \dots, T\}$  is the set of observations on subject  $i$  and similarly  $\{n_{it}\}$  the set of values of the nutrient level. Unconditionally this likelihood is  $\int_0^1 L_T(i, \lambda) f(\lambda) d\lambda$ , and the likelihood of all  $I$  subjects is the product

$$L_T = \prod_{i=1}^I \int_0^1 L_T(i, \lambda) f(\lambda) d\lambda. \quad (16.8)$$

Each evaluation of  $L_T$  therefore requires  $I$  numerical integrations, which can be performed in **R** by means of the function `integrate`, but slow the computation down considerably.

Incorporation of more than one random effect could proceed similarly, but would require the specification of the joint distribution of these effects, and the replacement of each of the one-dimensional integrals appearing in Equation (16.8) by a multiple integral. The evaluation of such multiple integrals would of course make the computations even more time-consuming.

### 16.7.3 Inclusion of covariates

In some applications — although not the application which motivated this study — there may be covariate information available that could

help to explain observed behaviour, e.g. dietary differences, or whether subjects are male or female. The important question then to be answered is whether such covariate information can efficiently be incorporated into the likelihood computation. The building-blocks of the likelihood are the transition probabilities and initial distribution of the latent process, and the probabilities  $\pi_1$  and  $\pi_2$  of the behaviour of interest in the two states. Any of these probabilities can be allowed to depend on covariates without greatly complicating the likelihood computation.

If we wished to introduce a (possibly time-dependent) covariate  $y_t$  into the probabilities  $\pi_i$ , here denoted  $\pi_i(y_t)$ , we could take  $\text{logit } \pi_1(y_t)$  to be  $a_1 + a_2 y_t$ , and similarly  $\text{logit } \pi_2(y_t)$ . The likelihood evaluation would then present no new challenges, although the extra parameters would of course tend to slow down the optimization. If the covariate  $y_t$  were instead thought to affect the transition probabilities, we could define  $\text{logit } \gamma_{11}(n_t, y_t)$  to be  $\alpha_0 + \alpha_1 n_t + \alpha_2 y_t$ , and similarly  $\text{logit } \gamma_{22}(n_t, y_t)$ .

## 16.8 Other modifications or extensions

Other potentially useful extensions are to increase the number of latent states above two, and to change the nature of the state-dependent distribution, e.g. to allow for more than two behaviour categories or for a continuous behaviour variable.

### 16.8.1 Increasing the number of states

If more than two latent states are required in the model, this can be accommodated, e.g. by using at Equation (16.2) a higher-dimensional analogue of the logit transform. Any such increase potentially brings with it a large increase in the number of parameters, however; if there are  $m$  states, there are  $m^2 - m$  transition probabilities to be specified, and it might be necessary to impose some structure on the transition probabilities in order to reduce the number of parameters.

### 16.8.2 Changing the nature of the state-dependent distribution

In this work the observations are binary, and therefore so are the state-dependent distributions, i.e. the conditional distributions of an observation given the underlying state. But there might well be more than two behaviour categories in some series of observations, or the observations might be continuous. That would require the use, in the likelihood computation, of a different kind of distribution from the binary distribution used here. That is a simple matter, and indeed that flexibility at observation level is one of the advantages of HM or similar models; almost any

kind of data can be accommodated at observation level without greatly complicating the likelihood computation. But since here the observations feed back into the nutrient level  $N_t$ , the feedback mechanism would need to be correspondingly modified.

## 16.9 Application to caterpillar feeding behaviour

### 16.9.1 Data description and preliminary analysis

The model was applied to sequences of observations of eight final-instar *Helicoverpa armigera* caterpillars, collected in an experiment designed to quantify developmental changes in the pattern of feeding (Raubenheimer and Barton Browne, 2000). The caterpillars were observed continuously for 1132 minutes, during which time they were scanned at one-minute intervals and scored as either feeding or not feeding. In order to isolate developmental changes from environmental effects, individually-housed caterpillars were fed semi-synthetic foods of homogeneous and constant nutrient composition, and the recordings were made under conditions of constant temperature and lighting. The caterpillars were derived from a laboratory culture, and so had similar ancestry and developmental histories. Figure 16.1 displayed the data.

Some remarks can immediately be made. Firstly, despite the uniform conditions of the experiment, there is considerable between-subject variation, both as regards the density of feeding events and the apparent pattern thereof. The runlengths of the feeding events differ between subjects; see e.g. those of subjects 5 and 6. A striking feature is that subject 8 began feeding only 6.2 hours after the start of the experiment. Closer examination of the original recordings made on this subject revealed that it was initially eating its exuvium (moulted skin), a behaviour which has been demonstrated to be of nutritional significance (Mira, 2000). However, since the nutrient composition of the exuvium is very different from that of the synthetic foods, in what follows only subjects 1–7 were included in the analysis. Also noticeable is the fact that subject 3 stopped feeding more than two hours before the end of the experiment. This anomaly became apparent in the model-checking exercise described below.

### 16.9.2 Parameter estimates and model checking

The first step in the model-fitting process was to fit a model separately to each of the seven subjects, i.e. to estimate for each one the nine parameters  $\alpha_0$ ,  $\alpha_1$ ,  $\beta_0$ ,  $\beta_1$ ,  $\pi_1$ ,  $\pi_2$ ,  $\lambda$ ,  $n_0$  and  $\delta_1$ . In all cases  $\delta_1$  was estimated as zero, i.e. the subject started in state 2. A convincing explanation for this is that, until the post-moult skin has hardened, insects cannot use

Table 16.1 *Parameter estimates and log-likelihood: individual models for subjects 1–7, and mixed model with six common parameters and random effect for  $\lambda$ .*

subj.	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\lambda}$	$\hat{n}_0$	$-\log L$
1	5.807	−11.257	2.283	2.301	0.936	0.000	0.027	0.240	331.991
2	2.231	−5.284	−0.263	21.019	0.913	0.009	0.032	0.150	347.952
3	4.762	−10.124	2.900	15.912	0.794	0.004	0.080	0.740	225.166
4	2.274	−7.779	1.294	16.285	0.900	0.000	0.056	0.018	298.678
5	3.135	−7.271	1.682	10.908	0.911	0.006	0.097	1.000	332.510
6	3.080	−5.231	1.374	13.970	0.880	0.001	0.043	0.246	291.004
7	3.888	−9.057	0.617	13.341	0.976	0.003	0.054	0.375	315.188
									2142.488
mixed model	2.735	−5.328	2.127	7.083	0.919	0.003	$\hat{\mu}=0.055$ $\hat{\sigma}=0.051$		2230.090

their mouthparts and so behave as if they were sated. In what follows we shall take it that all subjects start in state 2, and shall not further treat  $\delta_1$  as a parameter requiring estimation.

Table 16.1 displays *inter alia* the parameter estimates for each of subjects 1–7, and the corresponding values of minus the log-likelihood.

We now use subject 1 as illustrative. Figure 16.3 displays the observed feeding behaviour, the underlying motivational state sequence inferred by means of the Viterbi algorithm, and the nutrient level. Figure 16.4 presents an enlarged version of the observed feeding behaviour and inferred motivational state.

A point to note from the figures is the close correspondence between the series of feeding bouts and the inferred states. However, as is demonstrated by Figure 16.4, feeding bouts were interspersed with brief periods of non-feeding which did not break the continuity of the inferred state. The model thus succeeded in the aim of delimiting states according to the probability distributions of behaviours, rather than the occurrence of behaviours *per se*. The nutrient level for subject 1 ranges from about 0.1 to 0.5; for some other subjects the lower bound can reach zero. The parameter  $\lambda$  determines the (exponential) rate at which the nutrient level diminishes in the absence of feeding. The associated half-life is given by  $\log(0.5)/\log(1 - \lambda)$ . Thus the estimated half-life for subject 1 is approximately 25 minutes.

Figure 16.5 displays the transition probabilities for subject 1 as a function of nutrient level. As expected,  $\hat{\gamma}_{11}$  decreases with increasing

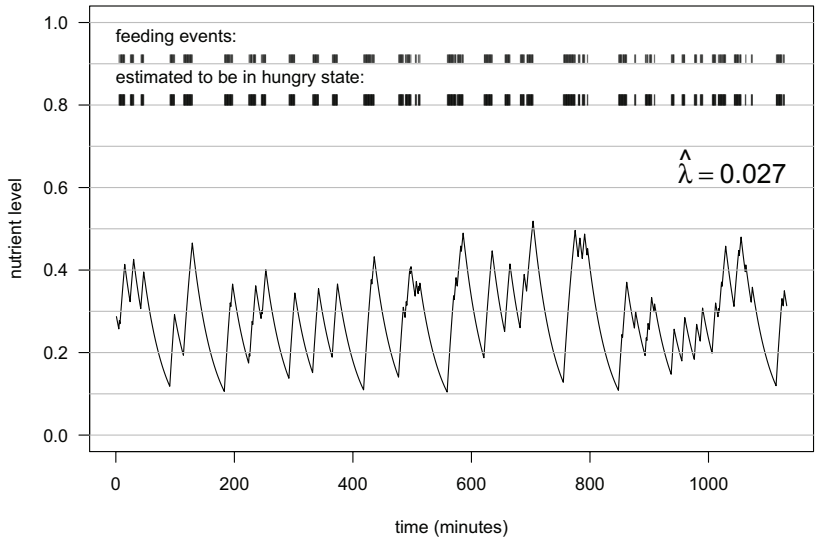


Figure 16.3 *Feeding behaviour, inferred motivational state and nutrient level for subject 1.*

nutrient level, and  $\hat{\gamma}_{22}$  increases, and this is true for all seven subjects. Note also that  $\hat{\pi}_2$ , the estimated probability of feeding when sated, is close to zero. In fact  $\hat{\pi}_2$  is less than 0.01 for all seven subjects, and less than 0.001 for two of them. The fact that these estimates are so close to the boundary of the parameter space has implications when one attempts to estimate standard errors. The standard errors of the parameters for subject 1 were estimated by the parametric bootstrap, but are not included here.

Table 16.2 gives runlength statistics, for subjects 1 to 7, of the observed feeding sequences and the estimated sequences of the state ‘hungry’. As expected, there are fewer runs for the latter (23% fewer on average) and the mean runlength is larger (45% on average), as is the standard deviation (20% on average).

In applying the model-checking technique of Section 16.5 to subjects 1–7, we were unable to reject  $H_0$  in six cases. Using the chi-squared approximation to the distribution of deviance differences, we obtained  $p$ -values ranging from 0.30 to 0.82 in these six cases. In the case of subject 3

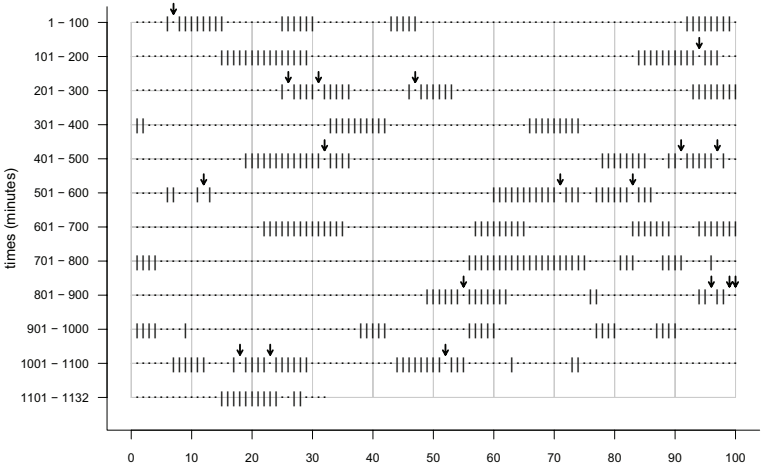


Figure 16.4 *Feeding behaviour of subject 1 (vertical lines), with arrows indicating times at which the subject was not feeding but was inferred to be in state 1, 'hungry'.*

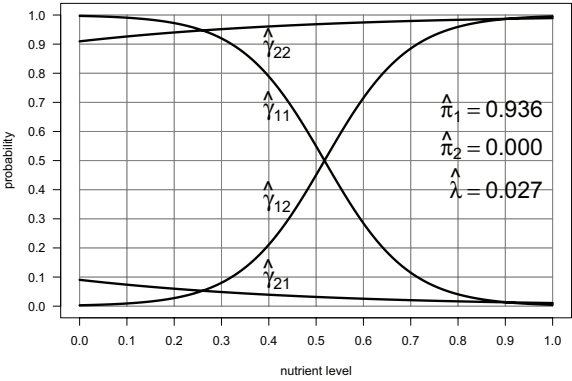


Figure 16.5 *Transition probabilities for subject 1.*

Table 16.2 *Runlength statistics. The seven columns are: subject no.; number of feeding runs, mean length of feeding runs, standard deviation of the length of feeding runs; number of (estimated) hungry runs, mean length of hungry runs, standard deviation of the length of hungry runs.*

subject	feeding runs			estimated hungry runs		
	number	mean	s.d.	number	mean	s.d.
1	58	5.4	4.1	41	8.1	4.9
2	67	3.0	2.3	53	3.9	2.8
3	41	3.1	2.1	22	6.7	2.4
4	57	2.6	1.5	51	3.0	1.7
5	65	2.8	1.6	54	3.5	2.0
6	51	4.1	2.8	35	6.4	3.8
7	57	4.1	2.4	52	4.6	2.7

we concluded that the model fitted is unsatisfactory ( $p = 0.051$ ). Except for subject 3, AIC would also select the hypothesized model.

Examining subject 3 more closely, we note that some of the parameter values are atypical; e.g. the probability of feeding when hungry is atypically low, only about 0.8. The data for subject 3 revealed the unusual feature that there were no feeding events after about time 1000, i.e. no feeding for more than two hours. This is clearly inconsistent with the earlier behaviour of this subject. This conclusion was reinforced by a plot of deviance residuals for subjects 1 and 3, along with spline smooths of these.

16.9.3 *Runlength distributions*

One of the key questions of biological interest that motivated this work was to assess the extent to which runs of feeding events differ from the runs in motivational state 1 (hungry), and similarly, runs of non-feeding events from runs in state 2 (sated). In an HMM the distributions of the runlengths in the two motivational states are geometric; that cannot be assumed to be the case here. Monte Carlo methods were used here to estimate the runlength distributions under the model.

For each of the seven subjects, a series of length 1 million was generated from the relevant fitted model, and the distribution of each of the four types of run estimated. Figure 16.6 displays plots and summary statistics of the four estimated distributions for subject 1.

The distribution of the feeding runlength clearly differs from that in the hungry state in the expected way. The probability of a runlength being one is almost twice as great for feeding runs as it is for hungry

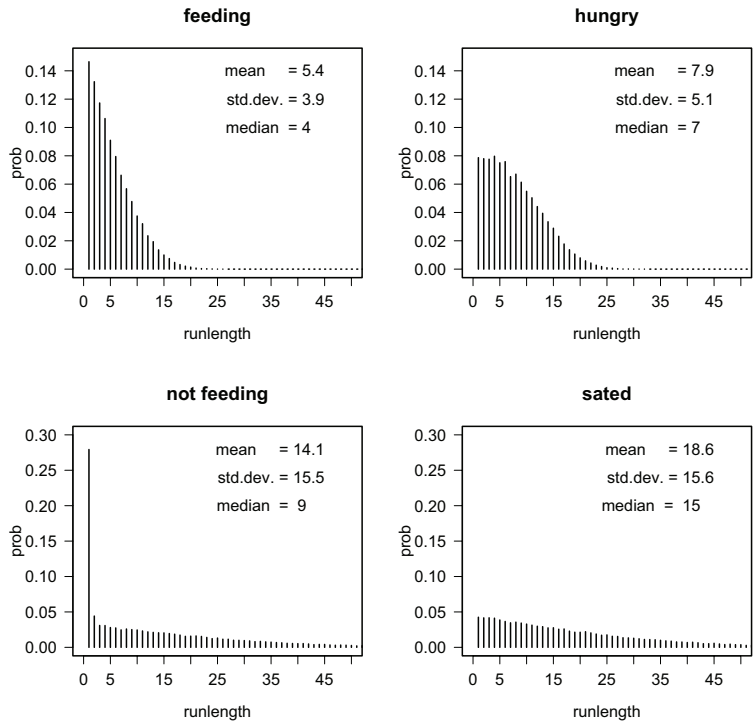


Figure 16.6 *Estimated runlength distributions for subject 1.*

Table 16.3 *Summary of models fitted jointly to the seven subjects. The number of parameters estimated is denoted by  $k$ . AIC selects the model with  $\pi_1$  and  $\pi_2$  common, and BIC the mixed model, which treats  $\lambda$  as a random effect.*

Model no.	Description	$-\log L$	$k$	AIC	BIC
1	no parameters common	2142.488	56	4396.976	4787.725
2	$\pi_1$ and $\pi_2$ common	2153.669	44	<b>4395.338</b>	4702.355
3	six parameters common	2219.229	20	4478.458	4618.011
4	seven parameters common	2265.510	14	4559.020	4656.707
5	mixed model (one random effect)	2230.090	15	4490.180	<b>4594.845</b>



runs. The mean and median for the latter are greater. The duration of a hungry spell is often longer than one would conclude if one took feeding and hunger as synonymous.

Unlike the distribution of the sated runs, that of the non-feeding runs has a marked peak at 1. This peak is attributable to hungry subjects often interrupting feeding for one time unit, rather than subjects being sated for a single time unit.

A further property worth noting is that the estimated distributions of hungry and sated runlengths do indeed have properties inconsistent with those of geometric distributions (on the positive integers). If  $\mu$  denotes the mean of such a distribution, the variance is  $\mu^2 - \mu$ , and the standard deviations corresponding to means of 7.9 and 18.6 would be 7.4 and 18.1; for hungry and sated runs the estimated s.d.s are, however, 5.1 and 15.6 respectively.

#### 16.9.4 Joint models for seven subjects

Five different models were considered for the observations on all seven subjects:

1. a model with no parameters common to the subjects;
2. a model with only  $\pi_1$  and  $\pi_2$  common to the subjects;
3. a model with the six parameters  $\alpha_0$ ,  $\alpha_1$ ,  $\beta_0$ ,  $\beta_1$ ,  $\pi_1$  and  $\pi_2$  common to the subjects;
4. a model with the seven parameters  $\alpha_0$ ,  $\alpha_1$ ,  $\beta_0$ ,  $\beta_1$ ,  $\pi_1$ ,  $\pi_2$  and  $\lambda$  common to the subjects;
5. a model which incorporated a random effect for the parameter  $\lambda$  and common values for the six parameters  $\alpha_0$ ,  $\alpha_1$ ,  $\beta_0$ ,  $\beta_1$ ,  $\pi_1$  and  $\pi_2$ . The model used for  $\lambda$ , which is bounded by 0 and 1, was a normal distribution (mean  $\mu$  and variance  $\sigma^2$ ) restricted to the interval  $[0,1]$ . This distribution was suggested by a kernel density estimate of  $\lambda$ -estimates from model 3.

In all cases the subjects were taken to be independent of each other and to have started in state 2. As there were only seven subjects, models with more than one random effect were not fitted because one cannot expect to identify a joint distribution of random effects from only seven observations. In the case of model 5, it was necessary to replace a small part of the **R** code by code written in C in order to speed up the computations.

Table 16.3 displays in each case the log-likelihood attained, the number of parameters estimated, AIC and BIC. The model selected by AIC is model 2, which has  $\pi_1 = 0.913$  and  $\pi_2 = 0.002$  for all subjects. The model selected by BIC is model 5, the mixed model. The parameter estimates for that model appear in Table 16.1, along with those of the

seven individual models. An interesting point to note is how much better the mixed model is than model 4. These two models differ only in their treatment of  $\lambda$ ; model 4 uses a single fixed value, and the mixed model uses a random effect.

### 16.10 Discussion

The application of HMMs to animal behaviour (MacDonald and Raubenheimer, 1995) has hitherto been limited to behaviours whose consequences do not readily alter motivational state.

The model here presented provides an extension which can allow for the important class of behaviours that are feedback-regulated (Toates, 1986). In the present example, the states ‘hungry’ and ‘sated’ as used above are of course an artefact of the model, and do not necessarily correspond to the accepted meanings of those terms. A different way of using the model described here would be to define state 1 as the state in which the probability of feeding is (say) 0.9, and similarly state 2 as that with probability (say) 0.1.

Irrespective of how the states are defined, the important point is that their delineation provides an objective means for exploring the physiological and environmental factors that determine the transitions between activities by animals. Such transitions are believed to play an important role in the evolution of behaviour (Sibly and McFarland, 1976), and the understanding of their causal factors is considered a central goal in the study of animal behaviour (Dewsbury, 1992).

The exponential filter — Equation (16.1) — used here for the nutrient level seems plausible. It is, however, by no means the only possibility, and no fundamental difficulty arises if this component of the model is changed. Ideally, the filter should reflect the manner in which feeding affects the motivational process, which might be expected to vary with a range of factors such as the nutrient composition of the foods, the recent feeding history of the animals, their state of health, etc. (Simpson, 1990). In our example, the same filter applied reasonably well across the experimental animals, probably because they were standardized for age and developmental history, and were studied in uniform laboratory conditions. Interestingly, however, a study of foraging by wild grasshoppers revealed that the patterns of feeding were no less regular than those observed in tightly controlled laboratory conditions (Raubenheimer and Bernays, 1993), suggesting that there might be some uniformity of such characteristics as the decay function even in more complex ecological conditions.

The models introduced here, or variants thereof, are potentially of much wider applicability than to feeding behaviour. They may be ap-

plied essentially unchanged to any binary behaviour thought to have a feedback effect on some underlying state, and with some modification of the feedback mechanism the restriction to binary observations could be removed.

In overview, by allowing for feedback-regulation we have extended the application of HM or similar models to a wider range of applications in the study of behaviour. We believe that these models hold potential for exploring the relationships among observed behaviours, the activities within which they occur, and the underlying causal factors. Accordingly, they provide a step towards a much-needed objective science of motivation (Kennedy, 1992).