# A LATENT–STATE MODEL FOR TIME SERIES OF ANIMAL BEHAVIOUR


Walter Zucchini
University of Göttingen

David Raubenheimer
University of Auckland

Iain MacDonald
University of Cape Town

# Outline

1. Background and objective
2. Two-state Bernoulli hidden Markov model
3. Structure of the proposed model
4. Caterpillar feeding experiment
   — Estimation and interpretation of the parameters
   — Decoding, Runlengths, Model checking
5. Extensions to the model
   — General state-dependent distributions
   — Incorportating covariates
   — Including random effects
6. Summary

# Background and objective

- Animal behaviourists study causal factor that determine behaviour, such as drinking, locomoting, grooming and feeding

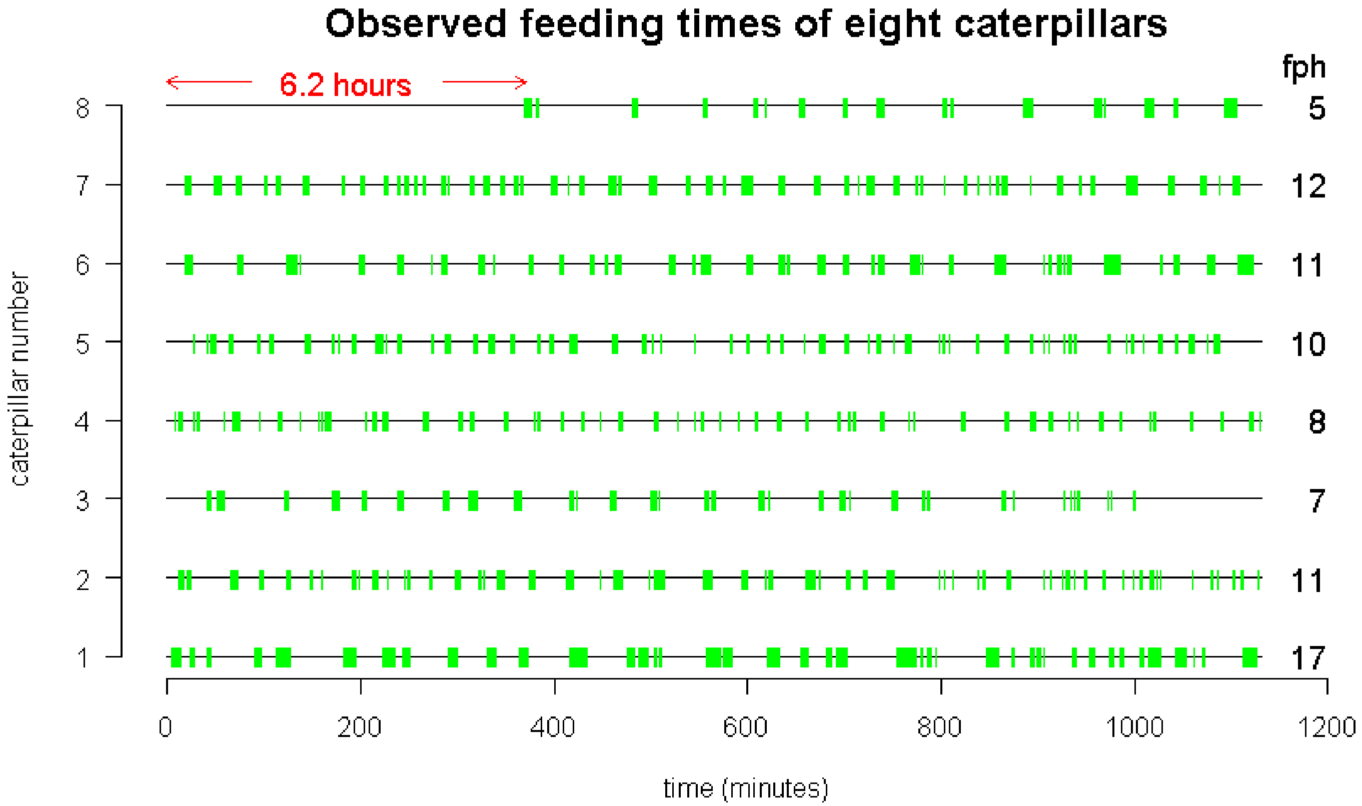
- Feeding behaviour results from the nervous system integrating information regarding

  — physiological factors: e.g. level of nutrients in the blood,
  — sensory inputs: e.g. perception of nutrients in food.

- The combined physiological and perceptual state of the animal is termed the **motivational state** (MacFarland, 1999).

**Raubenheimer and Barton Browne (2000)**
observed eight caterpillars *Helicoverpa armigera*
once per minute for 19 hours.

**Recordings:** feeding or not feeding
**Data:** 8 binary time series of length 1132
**Outlier:** One caterpillar was anomalous, and not modelled



**Assume** there are two motivational states — hungry and sated.

**Notation:**

$X_1, X_2, \ldots X_T$ sequence of observed (binary) feeding behaviour,

$C_1, C_2, \ldots C_T$ sequence of unobserved motivational states.

| behavioural state | (observed) | motivational state | (unobserved) |
|---|---|---|---|
| feeding | $X_t = 1$ | hungry | $C_t = 1$ |
| not feeding | $X_t = 0$ | sated | $C_t = 2$ |

The motivational state influences, but does not determine, behaviour.

A hungry animal doesn't always feed: $\pi_1 = \Pr(X_t = 1 | C_t = 1) < 1$

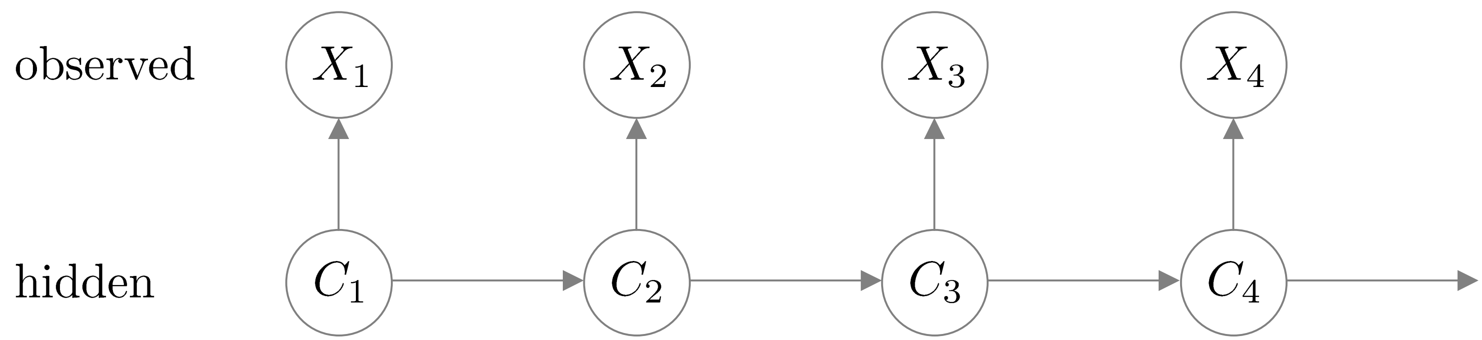
A sated animal sometimes feeds: $\pi_2 = \Pr(X_t = 1 | C_t = 2)) > 0$

**Objective:** Infer the motivational states from the observed behaviour.

**MacDonald and Raubenheimer (1995)**

used a Bernoulli–hidden Markov model (HMM) to describe this phenomon.

— motivation series: $C_1, C_2, \cdots$ homogeneous two-state Markov chain
— behaviour series: $X_1, X_2, \cdots$ mixture of two Bernoulli distributions
— assumption: conditional independence




**Definition of a HMM**

Notation: $X^{(t)}$ denotes the history up to time $t$, i.e. $\{X_t, X_{t-1}, \cdots, X_1\}$.

$$\begin{aligned} \Pr(C_t \,|\, C^{(t-1)}) &= \Pr(C_t \,|\, C_{t-1}) && \text{Markov property} \\ \Pr(X_t \,|\, X^{(t-1)}, C^{(t)}) &= \Pr(X_t \,|\, C_t) && \text{Conditional independence} \end{aligned}$$

**Transition probability matrix** of the homogeneous Markov chain:

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix} = \begin{pmatrix} \Pr(C_{t+1}=1 \,|\, C_t=1) & \Pr(C_{t+1}=2 \,|\, C_t=1) \\ \Pr(C_{t+1}=1 \,|\, C_t=2) & \Pr(C_{t+1}=2 \,|\, C_t=2) \end{pmatrix}$$

Note that $\begin{cases} \gamma_{11}+\gamma_{12}=1 \\ \gamma_{21}+\gamma_{22}=1 \end{cases}$

**Initial state distribution**: $\boldsymbol{\delta} = (\delta_1 \;\; \delta_2)$

If the chain is also stationary: $\boldsymbol{\delta} = \frac{1}{\gamma_{12}+\gamma_{21}} (\gamma_{21} \;\; \gamma_{12})$

**State-dependent distributions** $\begin{cases} X_t \,|\, C_t=1 & \sim & \text{Bernoulli}(\pi_1) \\ X_t \,|\, C_t=2 & \sim & \text{Bernoulli}(\pi_2) \end{cases}$

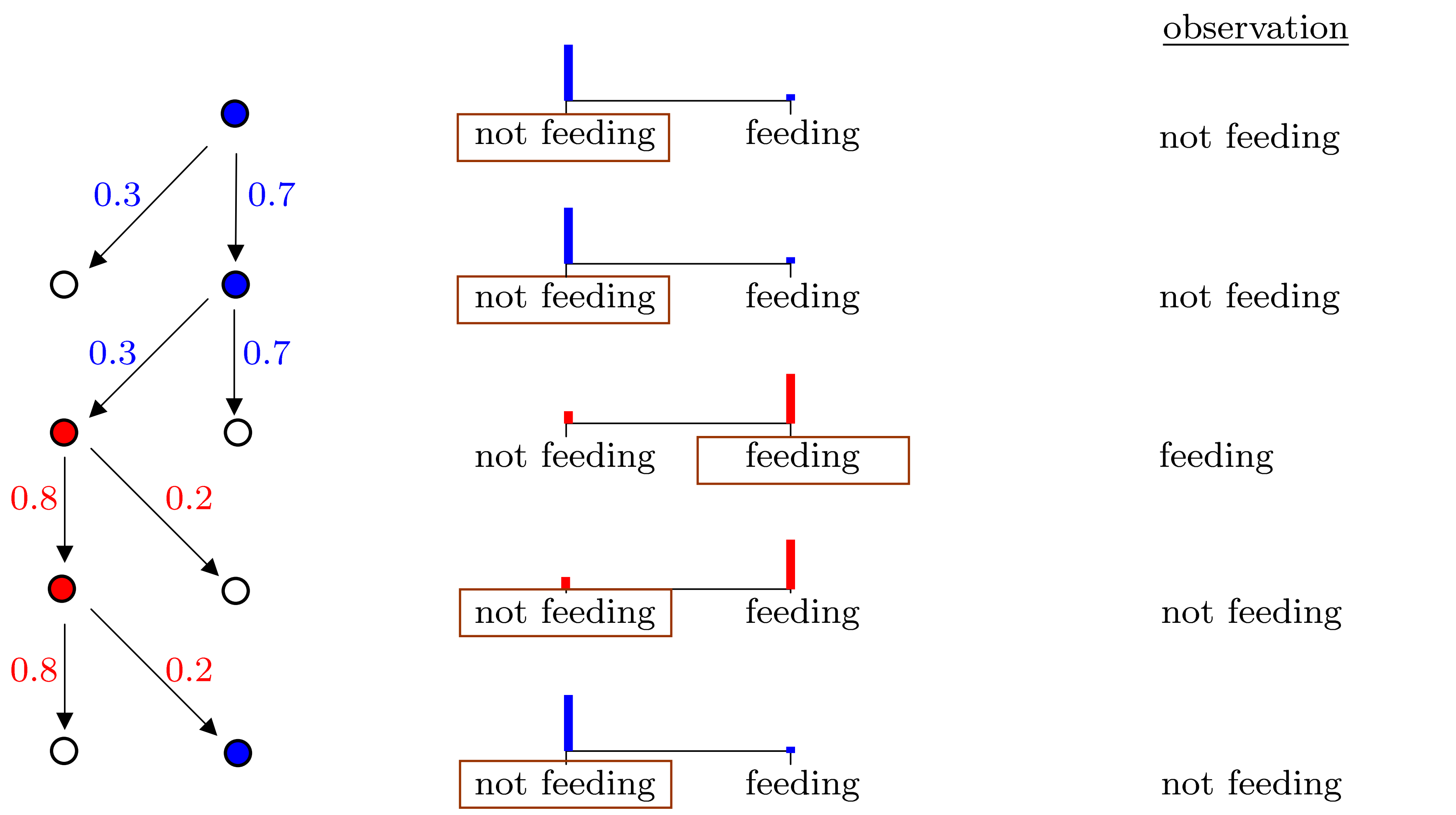
**Model parameters:**

State process (Markov chain): $\gamma_{11}$ $\gamma_{22}$ (and $\delta_1$ unless stationary)

State-dependent distributions: $\pi_1$ $\pi_2$

# Two-state Bernoulli HMM

| motivational state | | state-dependent distribution | transitional prob. matrix |
|---|---|---|---|
| state 1 | state 2 | $\pi_1 = P(\text{feed} \mid \text{state 1}) = 0.8$ | $\Gamma = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}$ |
| 0.6 | 0.4 | $\pi_2 = P(\text{feed} \mid \text{state 2}) = 0.1$ | |



hidden

| observation |
|---|
| not feeding |
| not feeding |
| feeding |
| not feeding |
| not feeding |

**The likelihood** of an homogeneous HMM:

$$\boldsymbol{L_T} = \boldsymbol{\delta}\, \boldsymbol{P}(x_1)\ \boldsymbol{\Gamma}\, \boldsymbol{P}(x_2)\ \boldsymbol{\Gamma}\, \boldsymbol{P}(x_3)\ \cdots \boldsymbol{\Gamma}\, \boldsymbol{P}(x_T)\ \boldsymbol{1'}$$

For a two–state Bernoulli–HMM:

$$\boldsymbol{P}(x) = \begin{pmatrix} {\pi_1}^x (1-\pi_1)^{1-x} & 0 \\ 0 & {\pi_2}^x (1-\pi_2)^{1-x} \end{pmatrix}$$

**Parameter estimation** via:
— an EM algorithm (Baum-Welch algorithm),
— direct numerical maximization (e.g. `nlm` or `optim` in `R`).

**Global decoding:** estimating the most likely motivational state sequence.

Maximize, w.r.t. $c_1, c_2, \ldots, c_T$, the conditional probability:
$\Pr\left(C^{(T)} = c^{(T)} \,|\, X^{(T)} = x^{(T)}\right)$
This solved using a dynamic programming method, the Viterbi algorithm.

**So what's the problem then?**

1. The runlength distributions in each motivational state is (implicitly) assumed to be **geometric**. Alcroft *et al.* (2004) fitted a semi-Markov model to overcome this criticism.

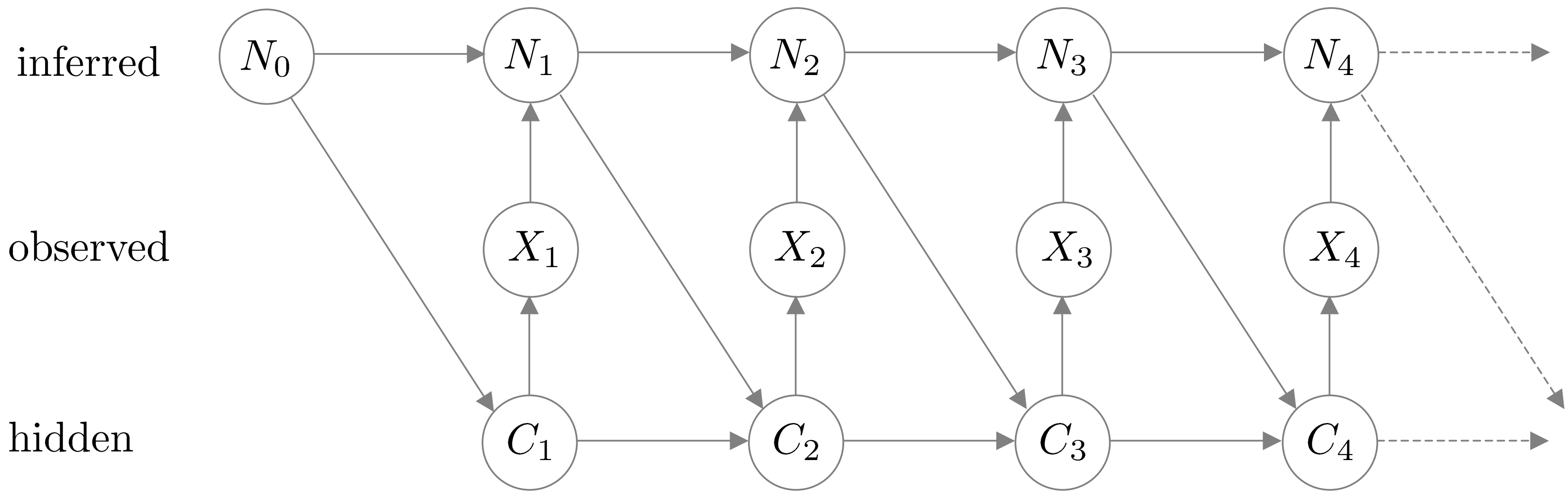
2. The model does not account for **feedback** from behaviour to motivation:
— feeding (eventually) leads to becoming sated;
— non-feeding (eventually) leads to hunger.

$$\begin{array}{ccc} \text{hungry} & \longrightarrow & \text{feed} \\ \uparrow & & \downarrow \\ \text{don't feed} & \longleftarrow & \text{sated} \end{array}$$

**Feedback loop:**
motivation → behaviour → motivation → behaviour

**Components of proposed model:**

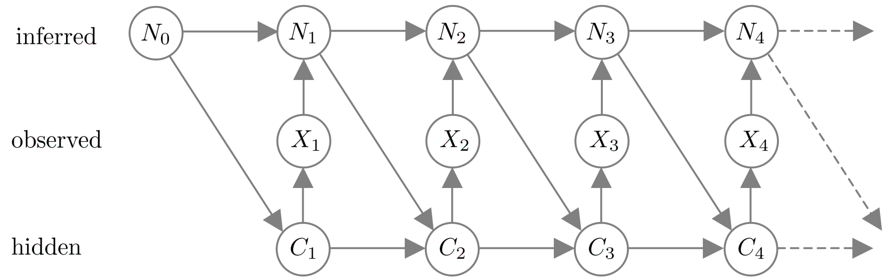
| | | |
|---|---|---|
| — motivation series: | $C_t$ | two-state process |
| — behaviour series: | $X_t$ | mixture of two Bernoulli distributions |
| — nutritional level: | $N_t$ | determined by feeding behaviour |
| — assumption: | | conditional independence |




**General idea:**

— The current state influences the feeding behaviour.
— Feeding behaviour determines the nutritional level.
— The nutritional level effects the probability of remaining in the current state.

**Model assumptions**




1. The motivational state at time $t$ depends **only** on the previous state and nutritional level.

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

2. Feeding behaviour at time $t$ depends only on motivational state.

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

3. Nutritional level is determined by feeding behaviour.

$$N_t = h(X^{(t)})$$

## State-transition behaviour

| current state | nutritional level | probable reaction | corresponding transition |
|---|---|---|---|
| hungry | low | remain hungry | $1 \rightarrow 1$ |
| hungry | high | become sated | $1 \rightarrow 2$ |
| sated | low | become hungry | $2 \rightarrow 1$ |
| sated | high | remain sated | $2 \rightarrow 2$ |

## A model for state transition behaviour

$$\mathbf{\Gamma}(n_t) = \begin{pmatrix} \gamma_{11}(n_t) & \gamma_{12}(n_t) \\ \gamma_{21}(n_t) & \gamma_{22}(n_t) \end{pmatrix}$$

The state transition probabilities, $\gamma_{ij}$ depend on $n_t$ as follows:

$$\gamma_{11}(n_t) = \frac{\exp(\alpha_0+\alpha_1 n_t)}{1+\exp(\alpha_0+\alpha_1 n_t)} \quad \text{i.e.} \quad \text{logit}(\gamma_{11}(n_t)) = \alpha_0 + \alpha_1 n_t$$

$$\gamma_{22}(n_t) = \frac{\exp(\beta_0+\beta_1 n_t)}{1+\exp(\beta_0+\beta_1 n_t)} \quad \text{i.e.} \quad \text{logit}(\gamma_{22}(n_t)) = \beta_0 + \beta_1 n_t$$

$\alpha_1 = \beta_1 = 0 \Longrightarrow$ no feedback from nutritional level to motivational state.

## A model for the nutritional level

The nutritional level is determined by the feeding behaviour as follows:

$N_t = \lambda X_t + (1-\lambda)N_{t-1}\,, \qquad t = 1, 2, \ldots, T.$ ($N_0$ is regarded as a parameter.)

$\lambda \in (0, 1)$ determines the rate of decay.

Contribution of one feeding episode has half-life $= \log(0.5)/\log(1-\lambda)$.

## Model parameters

| | |
|---|---|
| $\alpha_0$ $\alpha_1$ | determine how the nutritional level affects Pr(remaining hungry) |
| $\beta_0$ $\beta_1$ | determine how the nutritional level affects Pr(remaining sated) |
| $\pi_1$ $\pi_2$ | Pr(feed \| hungry) Pr(feed \| sated) |
| $\lambda$ | determines rate of nutrition depletion |
| $N_0$ | initial nutritional level |
| $\delta_1$ | $\Pr(C_1 = 1)$ |

**Likelihood** of the model:

$$\boldsymbol{L_T} = \boldsymbol{\delta}\, \boldsymbol{P}(x_1)\ \boldsymbol{\Gamma}\,(n_1)\, \boldsymbol{P}(x_2)\ \boldsymbol{\Gamma}\,(n_2)\, \boldsymbol{P}(x_3)\ \cdots \boldsymbol{\Gamma}\,(n_T)\, \boldsymbol{P}(x_T)\ \mathbf{1'}$$

$$\boldsymbol{P}(x) = \begin{pmatrix} {\pi_1}^x(1-\pi_1)^{1-x} & 0 \\ 0 & {\pi_2}^x(1-\pi_2)^{1-x} \end{pmatrix}$$

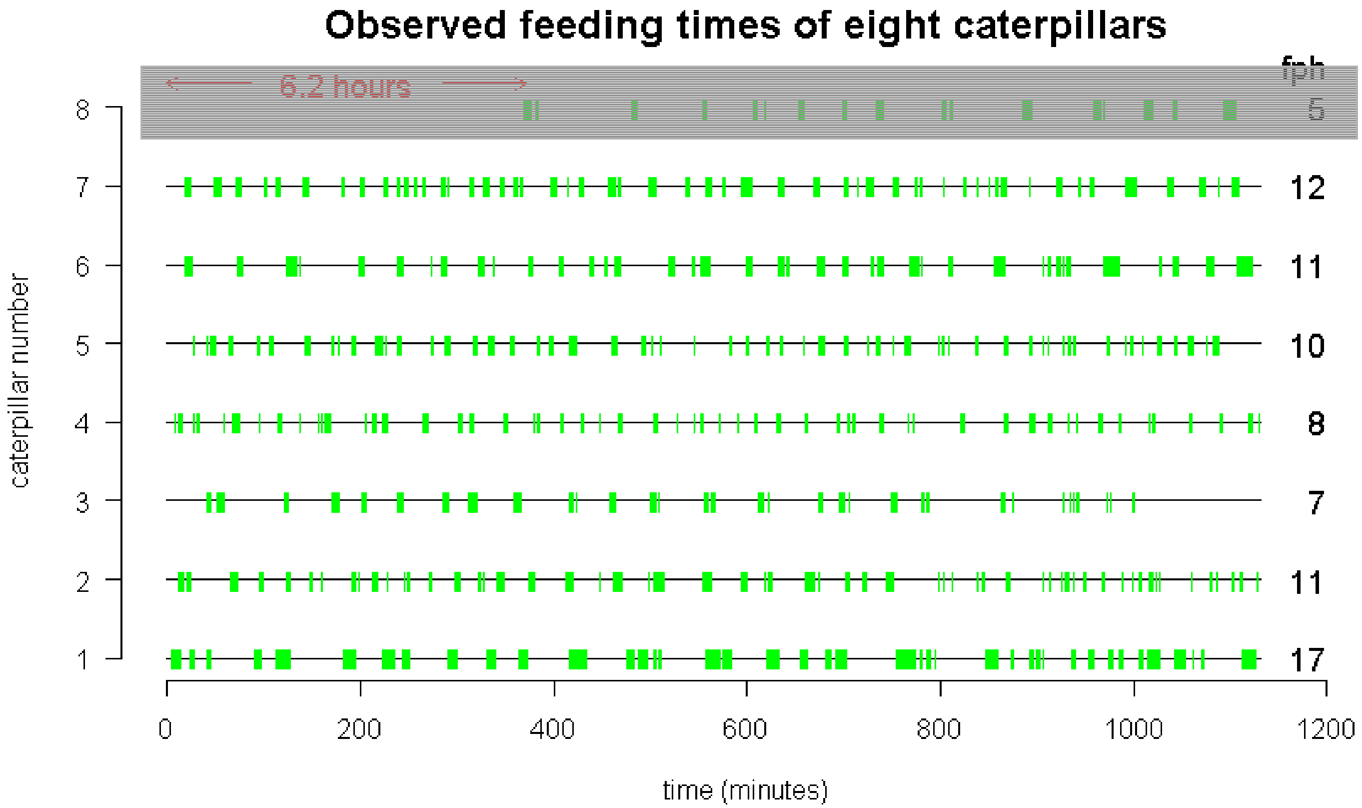
$$\boldsymbol{\Gamma}(n) = \begin{pmatrix} \frac{\exp(\alpha_0+\alpha_1 n)}{1+\exp(\alpha_0+\alpha_1 n)} & \frac{1}{1+\exp(\alpha_0+\alpha_1 n)} \\ \frac{1}{1+\exp(\beta_0+\beta_1 n)} & \frac{\exp(\beta_0+\beta_1 n)}{1+\exp(\beta_0+\beta_1 n)} \end{pmatrix}$$

**Parameter estimation** by direct numerical maximization (e.g. `nlm` in `R`)
(An EM algorithm would require numerical maximization in *each* M-step.)

**Global decoding:** estimating the most likely motivational state sequence.

The Viterbi algorithm is applicable.

**Back to the data**



**Parameter estimates for the seven caterpillars the seven**

| subj | $\widehat{\alpha}_0$ | $\widehat{\alpha}_1$ | $\widehat{\beta}_0$ | $\widehat{\beta}_1$ | $\widehat{\pi}_1$ | $\widehat{\pi}_2$ | $\widehat{\lambda}$ | $\widehat{n}_0$ | $-\log L$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 5.80 | -11.19 | 2.31 | 2.22 | 0.936 | 0.000 | 0.027 | 0.295 | 332.6 |
| 2 | 2.20 | -5.16 | -0.28 | 21.13 | 0.913 | 0.009 | 0.032 | 0.163 | 348.2 |
| 3 | 4.76 | -10.12 | 3.00 | 15.91 | 0.794 | 0.004 | 0.080 | 0.740 | 225.2 |
| 4 | 2.19 | -7.24 | 1.31 | 16.23 | 0.900 | 0.000 | 0.059 | 0.062 | 299.3 |
| 5 | 3.14 | -7.27 | 1.68 | 10.91 | 0.901 | 0.006 | 0.097 | 0.999 | 332.5 |
| 6 | 3.08 | -5.22 | 1.37 | 14.01 | 0.879 | 0.001 | 0.043 | 0.263 | 291.0 |
| 7 | 3.89 | -9.05 | 0.62 | 13.34 | 0.976 | 0.003 | 0.054 | 0.379 | 315.2 |

— All $\widehat{\alpha}_1 < 0$ and all $\widehat{\beta}_1 > 0$. (Expected).

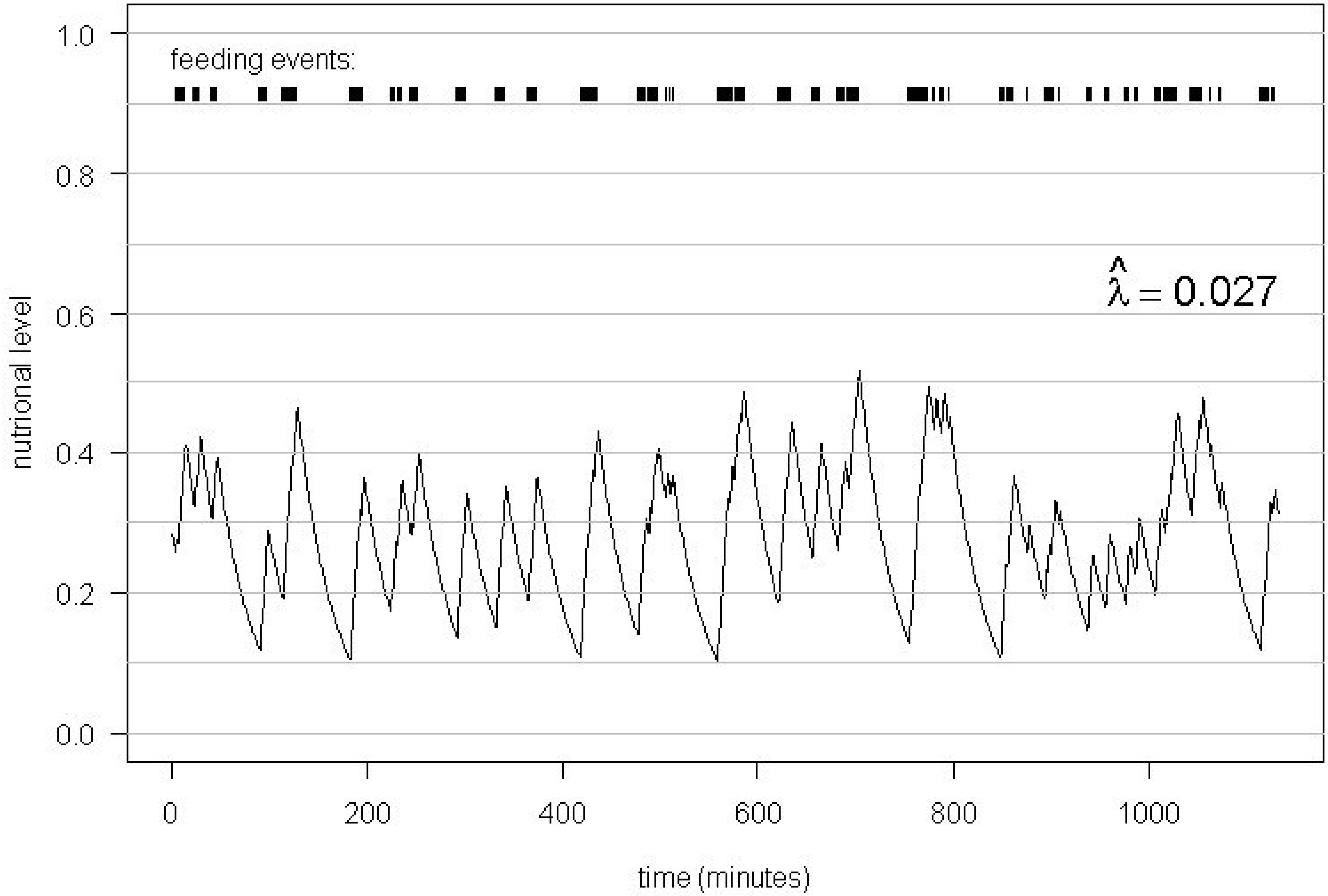
— $(\widehat{\alpha}_1, \widehat{\alpha}_2)$, and $(\widehat{\beta}_1, \widehat{\beta}_2)$, differ substantially between subjects, but the transition probabilities are not so different.

— All $\widehat{\pi}_1 \approx 1$ and all $\widehat{\pi}_2 \approx 0$.

— The estimates $\widehat{n}_0$ differ substantially. (Expected)

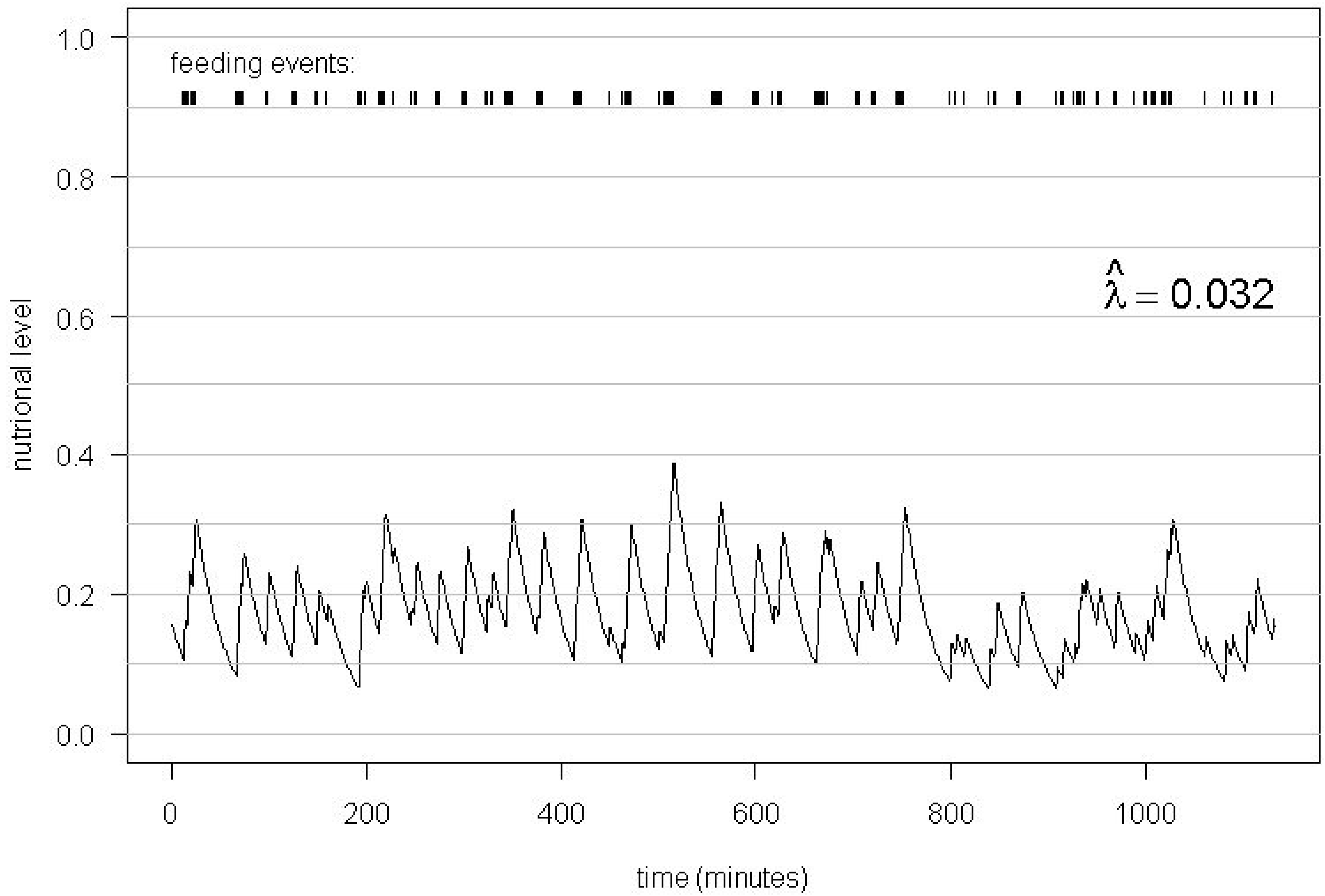
— The estimates $\widehat{\lambda}$ differ substantially. (Interesting)

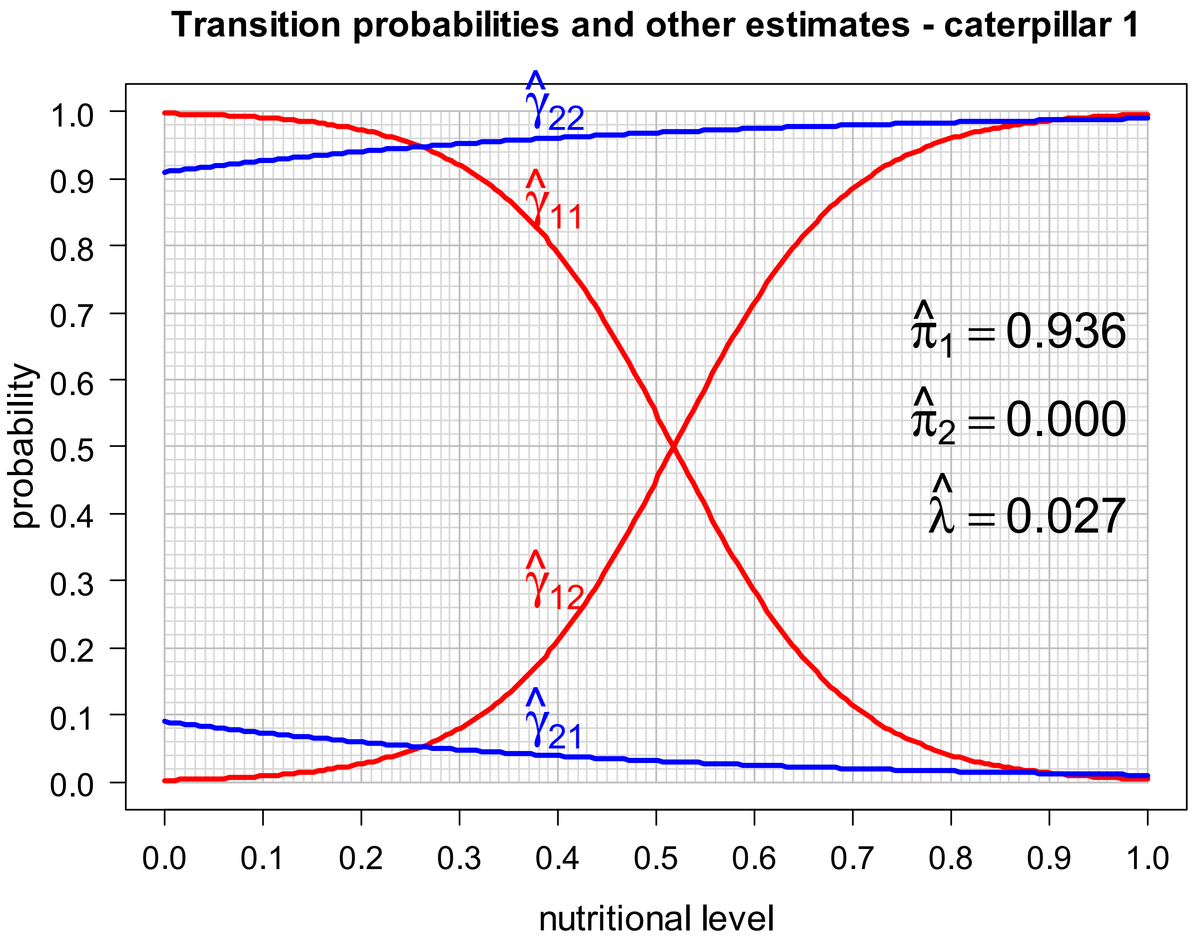
The nutritional level: Subject 1



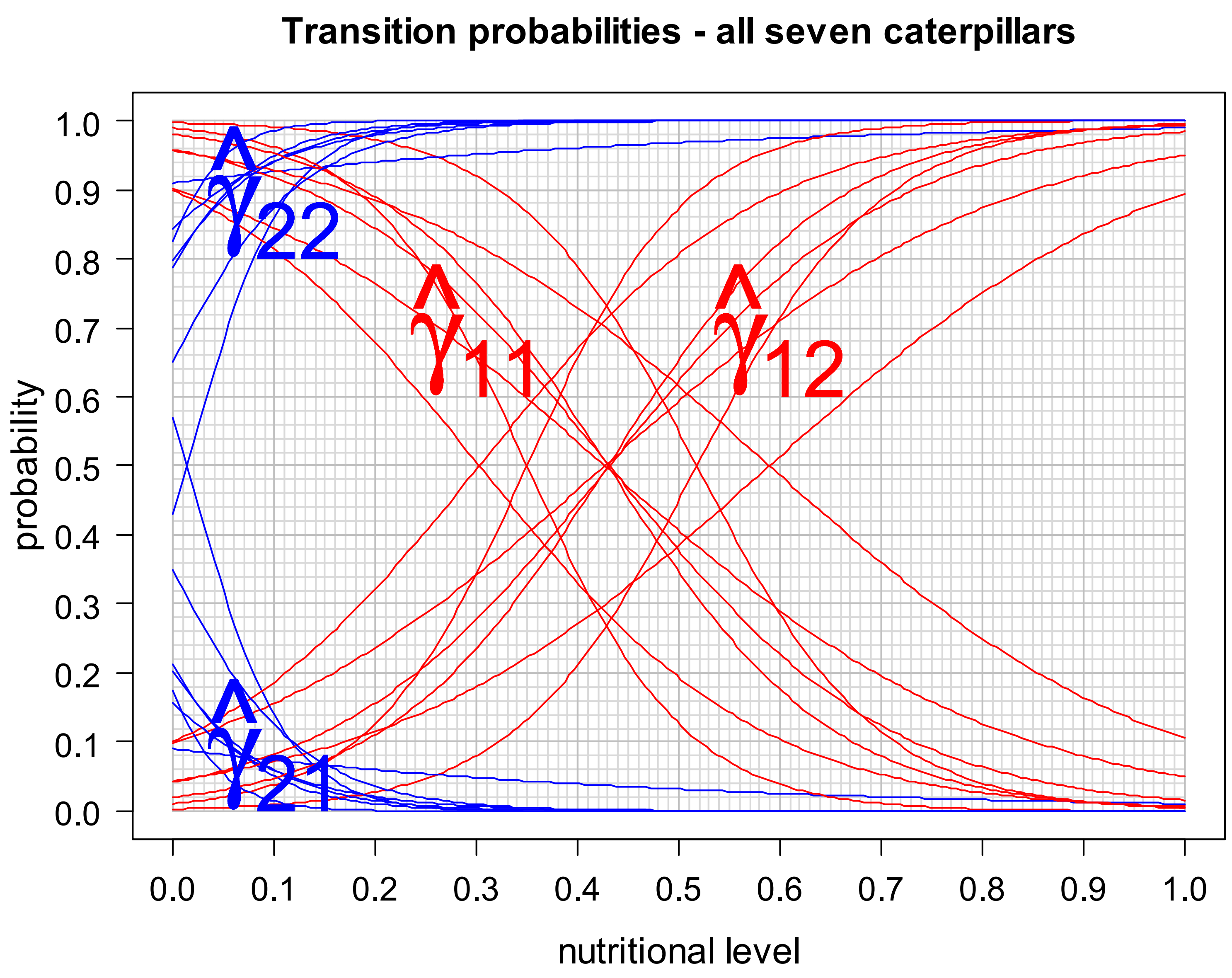

$$N_t = \lambda X_t + (1 - \lambda) N_{t-1}$$

The nutritional level: Subject 2



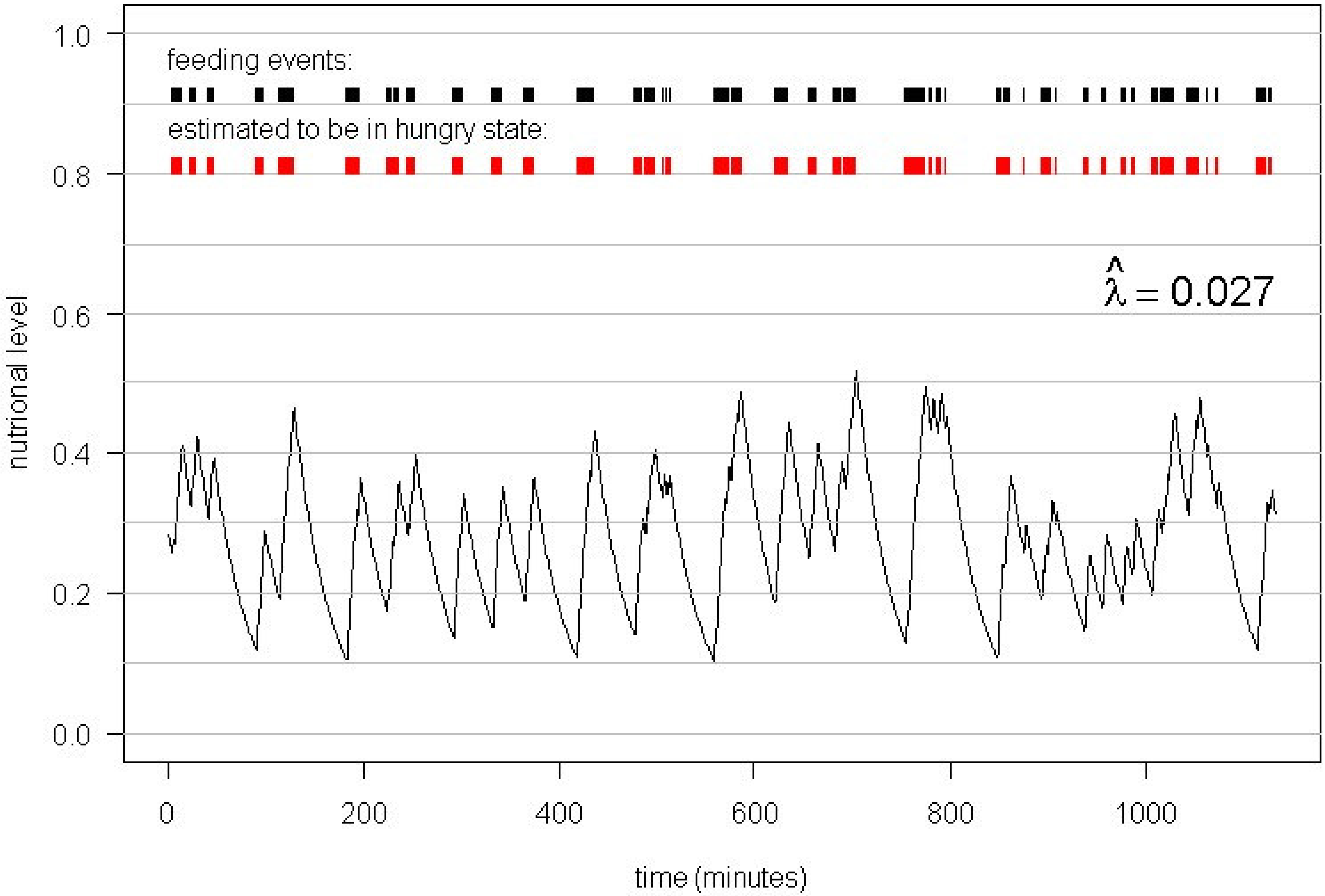



$$\mathbf{\Gamma}(n_t) = \begin{pmatrix} \gamma_{11}(n_t) & \gamma_{12}(n_t) \\ \gamma_{21}(n_t) & \gamma_{22}(n_t) \end{pmatrix}$$



Transition probabilities - all seven caterpillars
probability
nutritional level
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
$\hat{\gamma}_{22}$
$\hat{\gamma}_{11}$
$\hat{\gamma}_{12}$
$\hat{\gamma}_{21}$

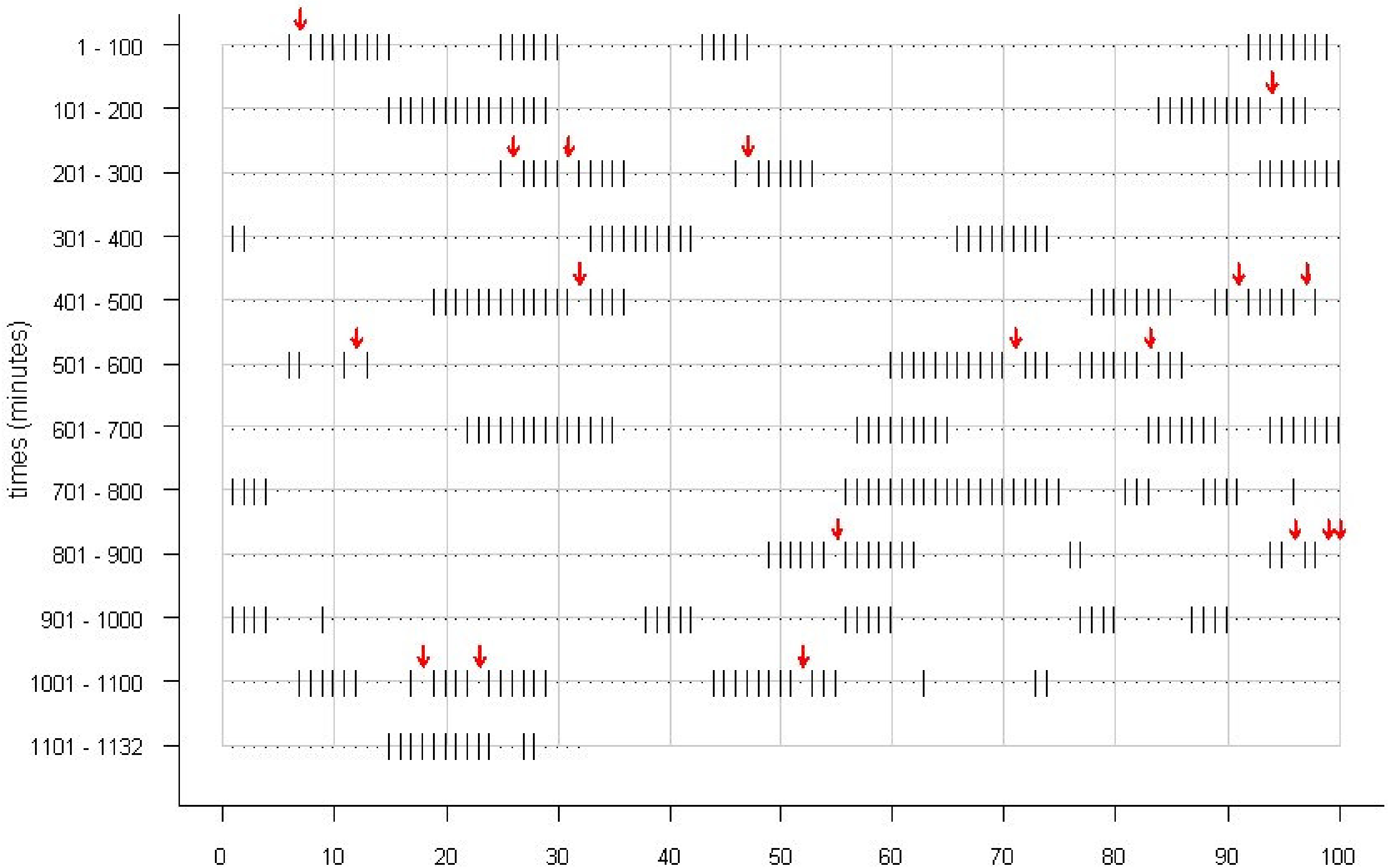
**Global decoding:** Subject 1



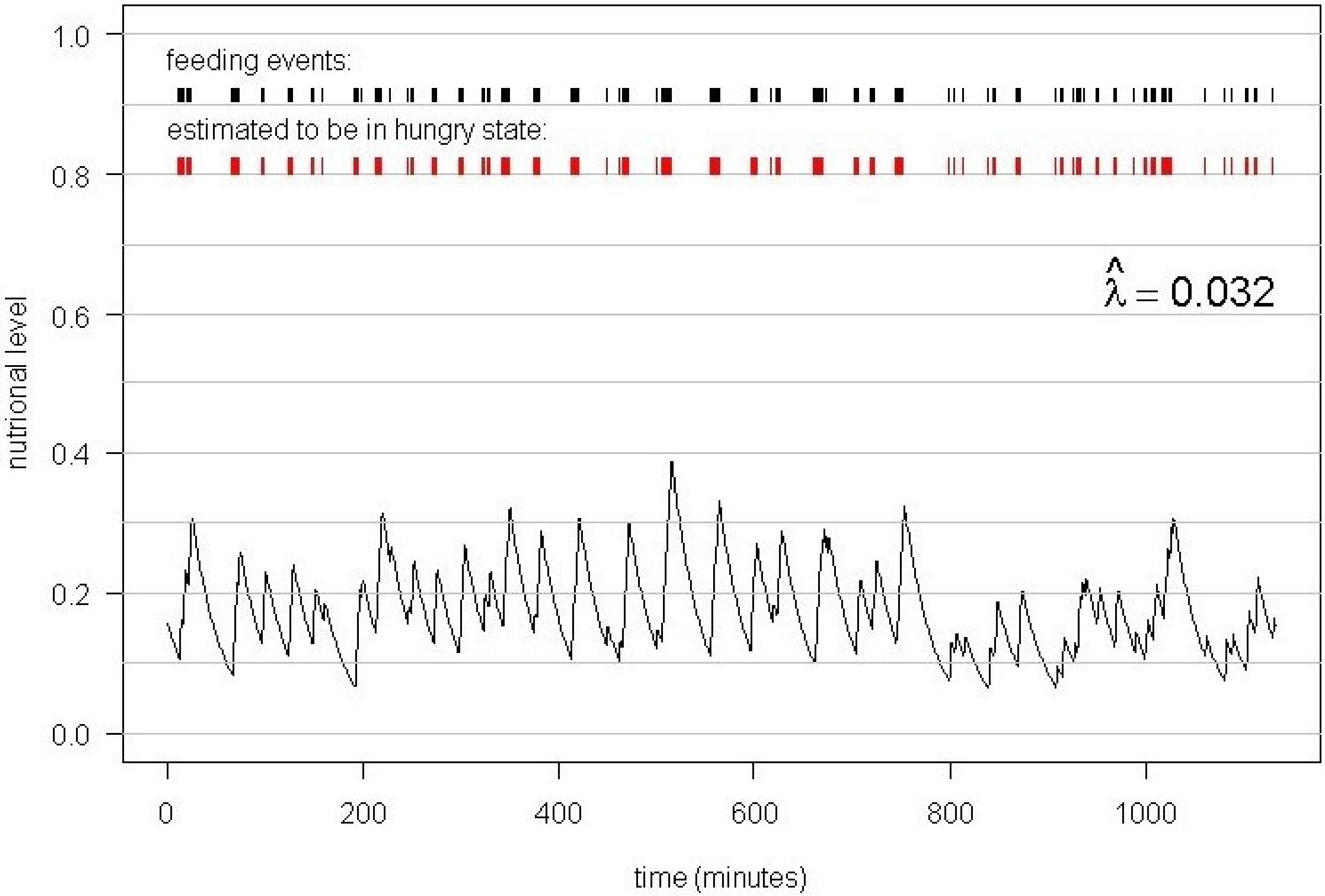
**Global decoding:** Subject 1

Down-arrows indicate non-feeding event while hungry,

Up-arrows indicate feeding event while sated.



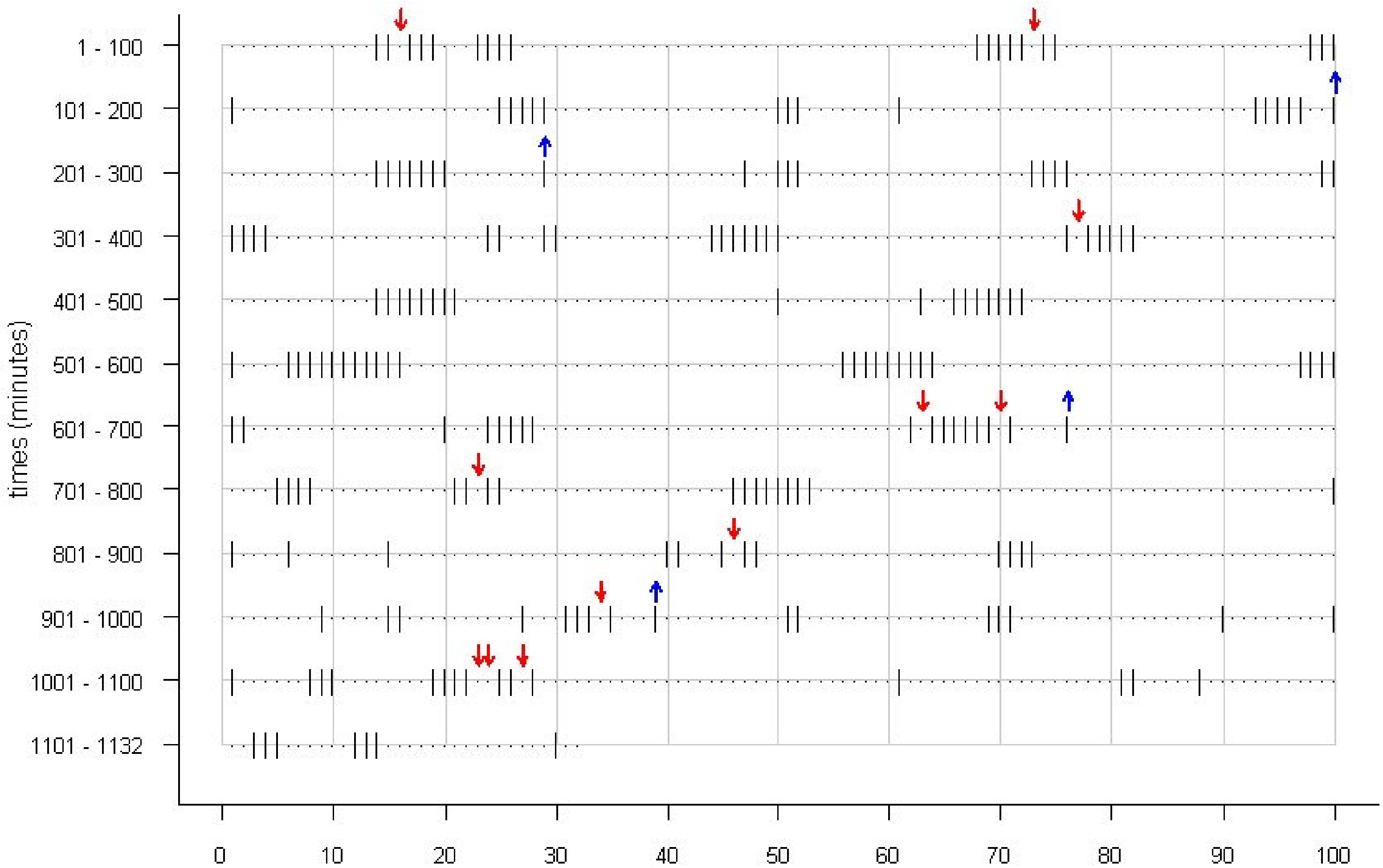
**Global decoding:** Subject 2



**Global decoding:** Subject 2

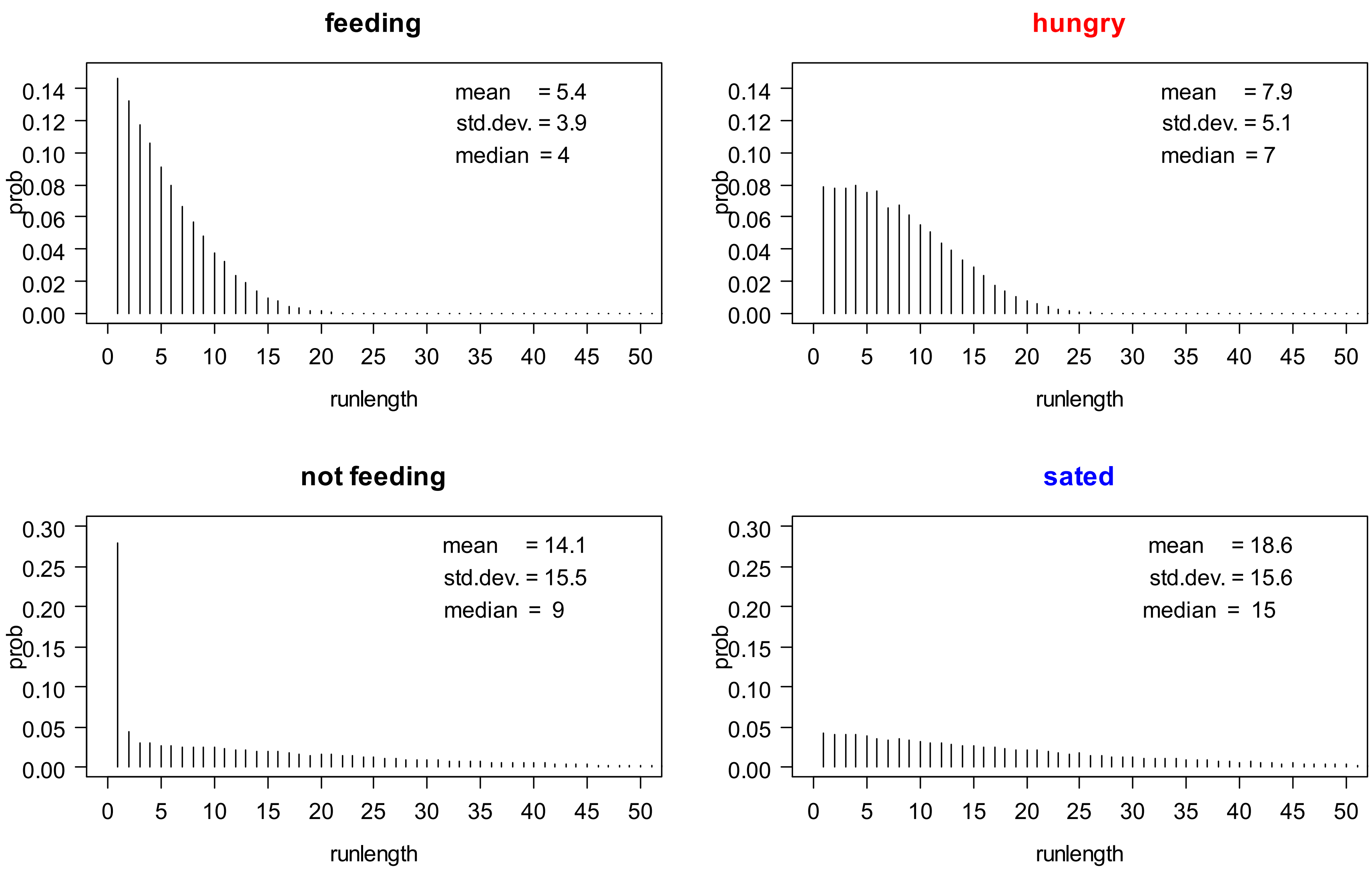
Down-arrows indicate non-feeding event while hungry,

Up-arrows indicate feeding event while sated.



## Runlength distributions for caterpillar 1

Notice the difference between feeding runs and hungry runs,
non-feeding runs and sated runs.



## Runlength statistics

| 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 |

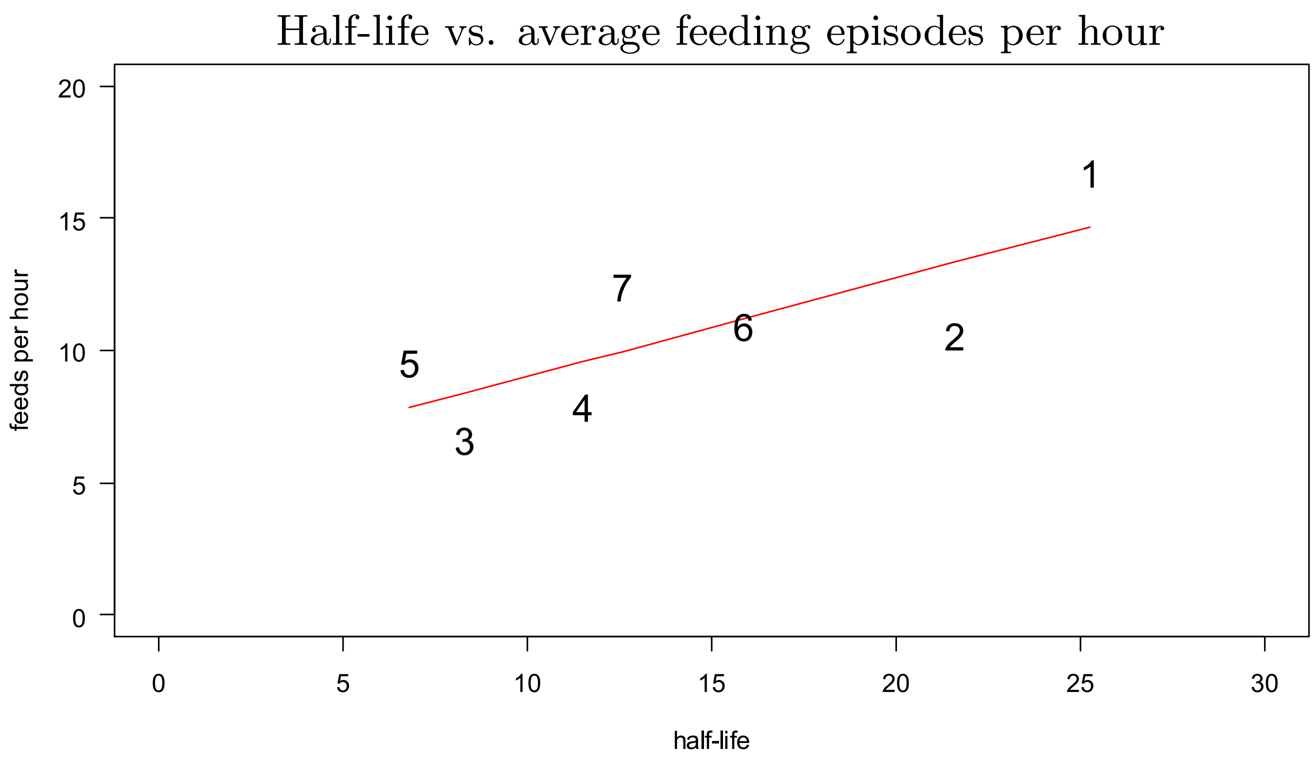
$$\text{Average}\left(\frac{\text{number of feeding runs}}{\text{number of hunger runs}}\right) = 1.35$$

$$\text{Average}\left(\frac{\text{mean feeding runlength}}{\text{mean hungry runlength}}\right) = 0.73$$

$$\text{Average}\left(\frac{\text{std. dev. feeding runlength}}{\text{std. dev. hungry runlength}}\right) = 0.83$$

## Estimated half-life

Half-life $= \log(0.5)/\log(1-\lambda)$
The time taken to halve the nutritional level when not feeding.




Half-life is related to the rate of feeding ($\widehat{\rho} = 0.77$)
The rate of feeding differs substantially between subjects.

**Estimating standard errors**

— either the "delta method" based on the estimated information matrix
— or parametric bootstrap (very computer intensive!)

**Model checking – 1. forecasts**

The forecast distribution: $\widehat{p}_t = \Pr(X_t = 1 \mid X^{(t-1)})$ is easy to compute. We test

$$\mathbf{H_0} : \boldsymbol{g(\mathrm{E}(x_t)) = g(\hat{p}_t)} \quad \text{vs.} \quad \mathbf{H_A} : \boldsymbol{g(\mathrm{E}(x_t)) = f(g(\hat{p}_t))},$$

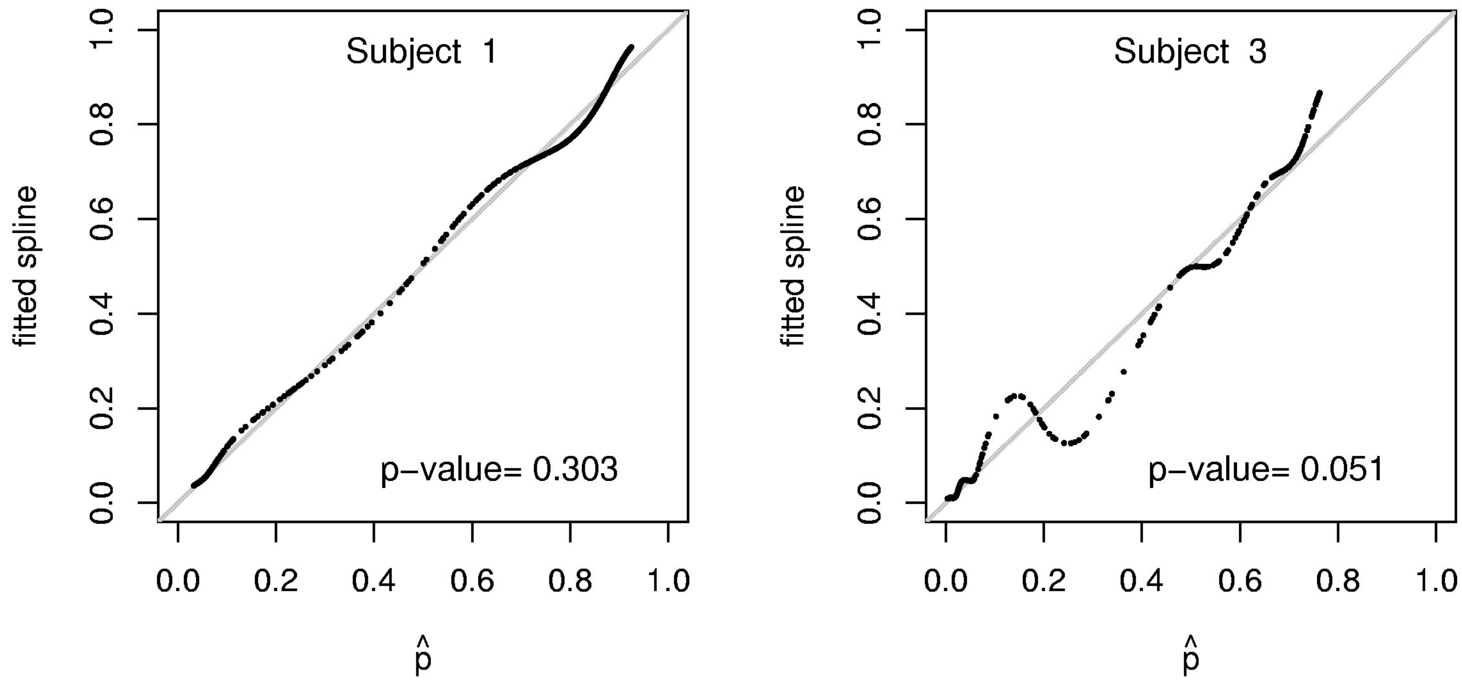
where $g$ is the logit function and $f$ a smoothing spline.

Departure of $f$ from the identity function constitutes evidence of a poor fit.

**Model checking – 2. deviance residuals**

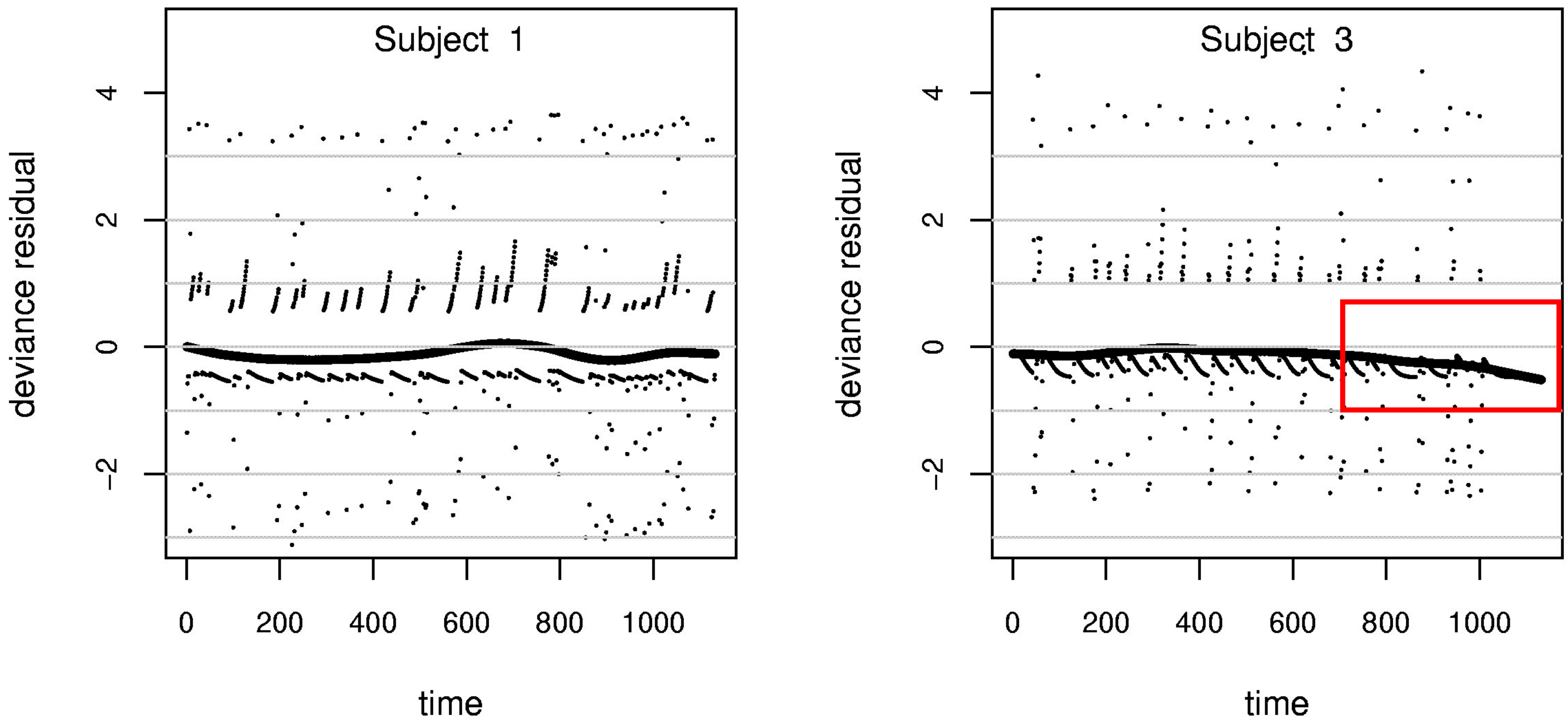
## Model checking – 1. forecasts

| subject | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $p$-value | 0.303 | 0.718 | 0.051 | 0.545 | 0.779 | 0.658 | 0.820 |



## Model checking – 2. deviance residuals

The solid line is a smooth of the deviance residuals.




Looking back at the data: subject 3 stopped feeding over the last 2 hours.

1. **The number of states can be increased to $m > 2$.**
   However, the number of parameters increases rapidly with increasing $m$.

2. **The definition of "nutritional level" can be changed.**
   The definition $N_t = \lambda X_t + (1 - \lambda) N_{t-1}$ is convenient but not essential.

3. **The state-dependent distribution can be changed.**
   — discrete-valued, continuous-valued, circular-valued distributions
   — multivariate, even mixed discrete-continuous, discrete-circular, etc.

   The likelihood remains of the form:

$$\boldsymbol{L_T} = \boldsymbol{\delta}\, \boldsymbol{P}(x_1)\ \boldsymbol{\Gamma}\,(n_1)\, \boldsymbol{P}(x_2)\ \boldsymbol{\Gamma}\,(n_2)\, \boldsymbol{P}(x_3)\ \cdots\ \boldsymbol{\Gamma}\,(n_T)\, \boldsymbol{P}(x_T)\ \boldsymbol{1'}$$

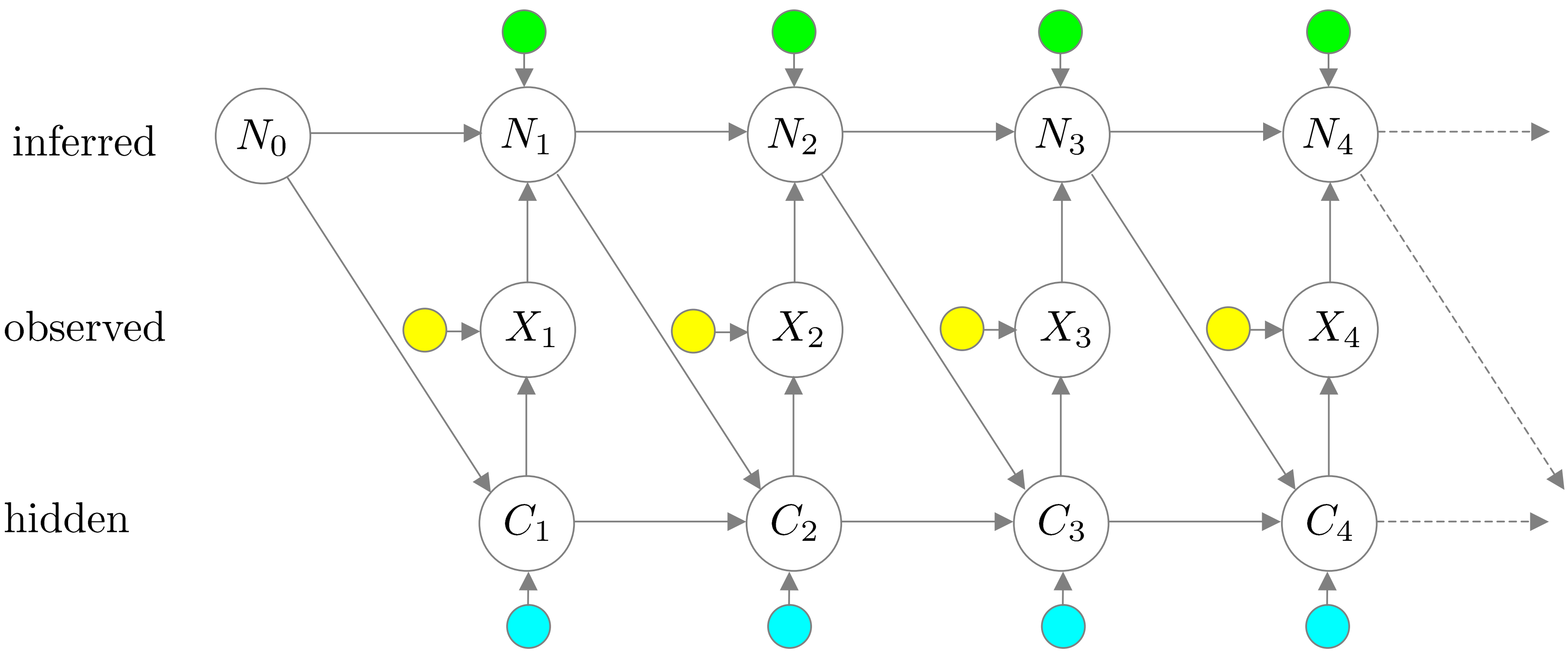
with $\boldsymbol{P}(x) = \begin{pmatrix} p_1(x) & 0 \\ 0 & p_2(x) \end{pmatrix}$

instead of $\boldsymbol{P}(x) = \begin{pmatrix} {\pi_1}^x (1 - \pi_1)^{1-x} & 0 \\ 0 & {\pi_2}^x (1 - \pi_2)^{1-x} \end{pmatrix}$

# Extensions of the model

**4. Covariate information can be included almost anywhere.**

- (green) — to influence the feedback state (nutritional level)
- (yellow) — to influence the state transition probabilities
- (cyan) — to influence the state-dependent distributions



**5. Mixed models for multiple time series.**

Analogous to the mixed hidden Markov models introduced by Altman (2007)

Some of the original parameters in the "caterpillar" model can be regarded as
**fixed effects** — the same for all caterpillars, or as
**random effects** — specific to individual caterpillars in a population.

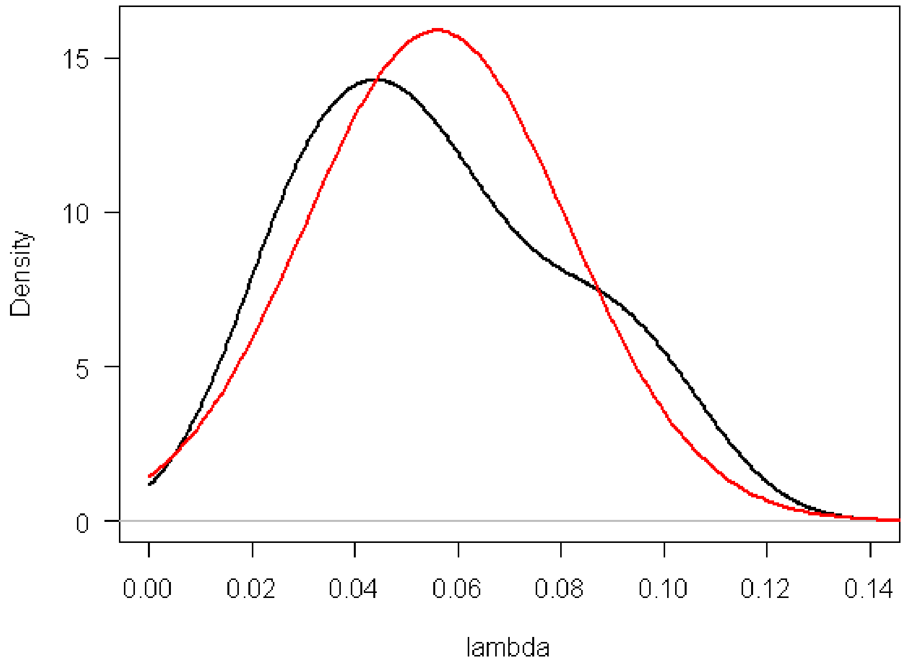
**Example based on the parameters of caterpillar model**

The assumptions below are only approximately applicable for our data. We make them to illustrate the technique.

| effect | fixed or random, i.e. the same for all individuals, or different? |
|---|---|
| $\alpha_0$ $\alpha_1$ | fixed |
| $\beta_0$ $\beta_1$ | fixed |
| $\pi_1$ $\pi_2$ | fixed |
| $\lambda$ | random |
| $N_0$ | random, but it's distribution is not of interest. |
| $\delta_1$ | random, but can be approximately determined via the other parameters, and will not be regarded as a free parameter. |

**Distribution for $\boldsymbol{\lambda}$**

Estimates of $f(\lambda)$;
truncated normal and truncated kernel




**Model:** $\lambda \sim$ (0,1)-truncated $N(\lambda; \mu, \sigma^2)$

$$f(\lambda; \mu, \sigma^2) = \frac{\phi(\frac{\lambda - \mu}{\sigma})}{\Phi(\frac{1-\mu}{\sigma}) - \Phi(\frac{0-\mu}{\sigma})}\ , \qquad \lambda \in (0, 1)$$

**Notation**

For $T$ observations on each of $I$ subjects, let

$x_{it}$ be the observation on subject $i$ at time $t$

$n_{it}$ be the nutrition level of subject $i$ at time $t$

**The likelihood**

$$\boldsymbol{L} = (\alpha_0,\, \alpha_1\, \beta_0,\, \beta_1,\, \pi_1\, \pi_2,\, N_0,\, \mu, \sigma^2;\ x_{it}, i = 1, 2, \cdots, I,\, t = 1, 2, \cdots T)$$

$$= \prod_{i=1}^{I} \int_0^1 \underbrace{(\boldsymbol{\delta}\, \boldsymbol{P}(x_{i1})\ \boldsymbol{\Gamma}\,(n_{i1})\, \boldsymbol{P}(x_{i2})\ \cdots \boldsymbol{\Gamma}\,(n_{iT})\, \boldsymbol{P}(x_{iT})\ \mathbf{1'})}_{\text{previous likelihood for subject i}}\ f(\lambda; \mu, \sigma^2)\ d\lambda$$

The likelihood can be maximized numerically with respect to the parameters.

— The numerical integration at each iteration makes this slow.

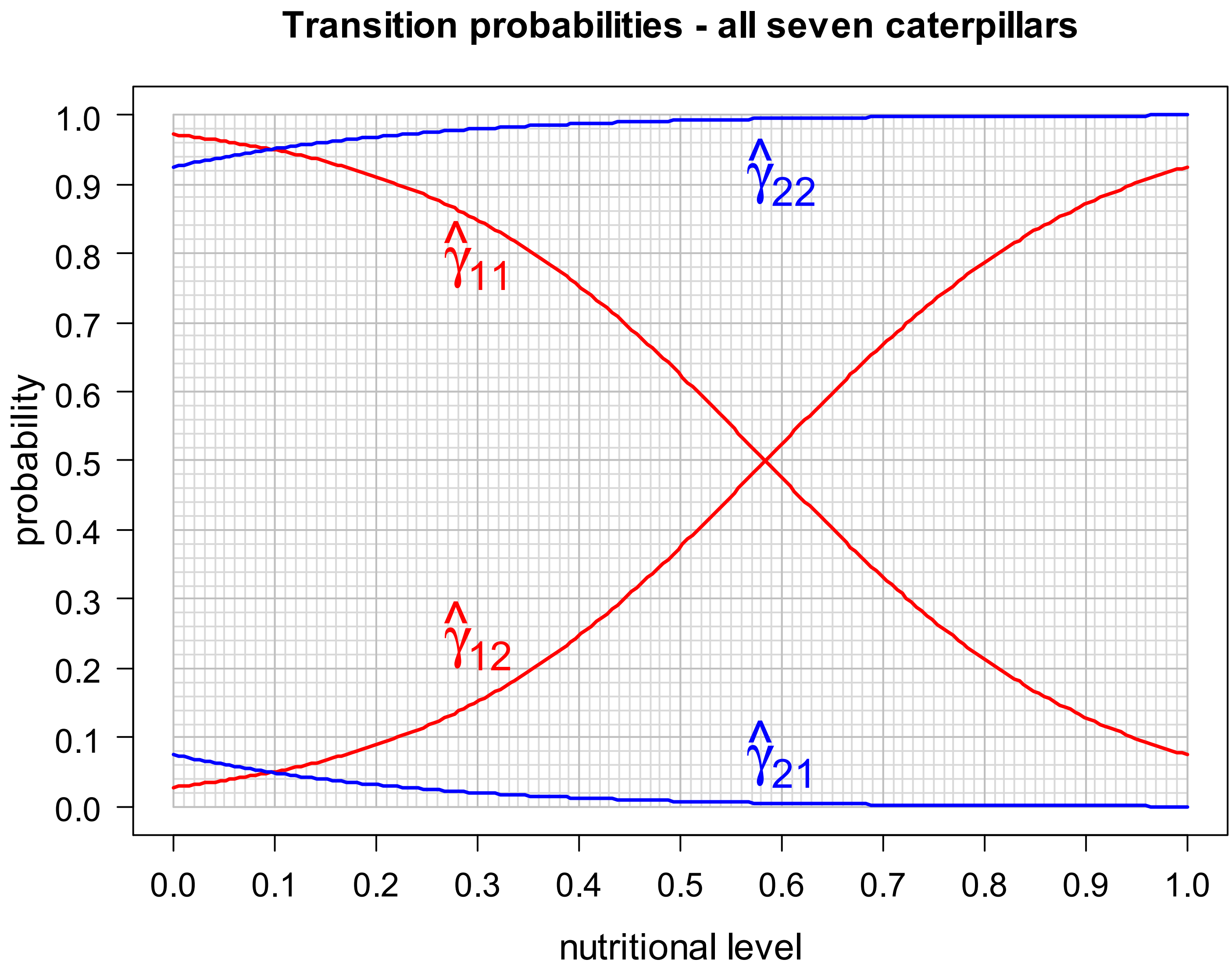
— Parameter constraints need to be respected, e.g. by reparameterization.

— Rescaling is needed to avoid numerical underflow.

**Parameter estimates for the original and the mixed model (brown).**

| subj | $\widehat{\alpha}_0$ | $\widehat{\alpha}_1$ | $\widehat{\beta}_0$ | $\widehat{\beta}_1$ | $\widehat{\pi}_1$ | $\widehat{\pi}_2$ | $\widehat{n}_0$ | $\widehat{\mathbf{n}}_\mathbf{0}$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 5.80 | -11.19 | 2.31 | 2.22 | 0.936 | 0.000 | 0.295 | **0.101** |
| 2 | 2.20 | -5.16 | -0.28 | 21.13 | 0.913 | 0.009 | 0.163 | **0.317** |
| 3 | 4.76 | -10.12 | 3.00 | 15.91 | 0.794 | 0.004 | 0.740 | **0.999** |
| 4 | 2.19 | -7.24 | 1.31 | 16.23 | 0.900 | 0.000 | 0.062 | **0.488** |
| 5 | 3.14 | -7.27 | 1.68 | 10.91 | 0.901 | 0.006 | 0.999 | **0.996** |
| 6 | 3.08 | -5.22 | 1.37 | 14.01 | 0.879 | 0.001 | 0.263 | **0.381** |
| 7 | 3.89 | -9.05 | 0.62 | 13.34 | 0.976 | 0.003 | 0.379 | **0.698** |
| **1-7** | **3.53** | **-6.04** | **2.50** | **4.55** | **0.921** | **0.006** | | |

Parameter estimates for the random effect : $\widehat{\boldsymbol{\mu}} = \mathbf{0.094}$ $\widehat{\boldsymbol{\sigma}} = \mathbf{0.058}$


Transition probabilities - all seven caterpillars
probability
nutritional level
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
$\hat{\gamma}_{11}$
$\hat{\gamma}_{12}$
$\hat{\gamma}_{21}$
$\hat{\gamma}_{22}$

## Model selection criteria

The estimates of the mixed model *look* reasonable *but* the model fits worse than the original ”full model”.

| Model | number of parameters | Akaike Information criterion |
|---|---|---|
| Full | 56 | 4400 |
| (*)Common $\pi_1, \pi_2$ | 44 | 4398 |
| Mixed effects | 15 | 4499 |

*The model with common values of $\pi_1, \pi_2$ for all subjects (and everything else different) achieved the best AIC of the models investigated.

Models with more than a single random effect were not investigated.
They take too long to fit (in **R**)!

**Positive aspects**

- The proposed models generalize the class of **hidden Markov models**, in that they allow for **feedback behaviour**.
- Like HMMs they are satisfyingly flexible.
- They differ from the class of **Markov switching models**, which are applied to model econometric time series and financial time series.

**Still needed:**

- Asymptotic properties of estimators need to be established.
- More efficent methods for estimating standard errors.
- More efficient methods for fitting mixed models. (It took 8 hours to fit the model with **one random effect** using **R**.)

## References

Allcroft, D.J., Tolkamp, B.J., Glasbey, C.A. and Kyriazakis, I. (2004), The importance of 'memory' in statistical models for animal feeding behaviour. *Behavioural Processes* **67**, 99–109.

Altmann, R.M. (2007), Mixed Hidden Markov Models: An Extension of the Hidden Markov Model to the Longitudinal Data Setting. *Journal of the American Statistical Association* **102 (477)**, 201–210.

MacDonald, I.L. and Raubenheimer, D. (1995), Hidden Marko,v models and animal behaviour. *Biometrical Journal* **37**, 701–712.

Raubenheimer, D. and Barton Browne, L. (2000), Developmental changes in the patterns of feeding in fourth- and fifth-instar *Helicoverpa armigera* caterpillars. *Physiological Entomology* **25**, 390–399.

## Other reading

Lange, K. (2002), *Mathematical and Statistical Methods for Genetic Analysis*, second edition. Springer, New York.

Langton, S.D., Collett, D. and Sibly, R.M. (1995), Splitting behaviour into bouts; a maximum likelihood approach. *Behaviour* **132**, 781–799.

MacDonald, I.L. and Zucchini, W. (1997), *Hidden Markov and Other Models for Discrete-valued Time Series*. Chapman & Hall, London.

Mira, A. (2000), Exuviae eating: a nitrogen meal? *Journal of Insect Physiology* **46**, 605–610.

Tolkamp, B.J., Allcroft, D.J., Austin, E.J., Nielsen, B.L. and Kyriazakis, I. (1998), Satiety splits feeding behaviour into bouts. *Journal of Theoretical Biology* **194**, 235–250.

Tolkamp, B.J. and Kyriazakis, I. (1999), To split behaviour into bouts, log-transform the intervals. *Animal Behaviour* **57**, 807–817.

Yeates, M.P., Tolkamp, B.J., Allcroft, D.J. and Kyriazakis, I. (2001). The use of mixture distribution models to determine bout criteria for analysis of animal behaviour. *Journal of Theoretical Biology* **213**, 413–425.