# Ensemble behaviour in population processes with applications to ecological systems

Phil Pollett

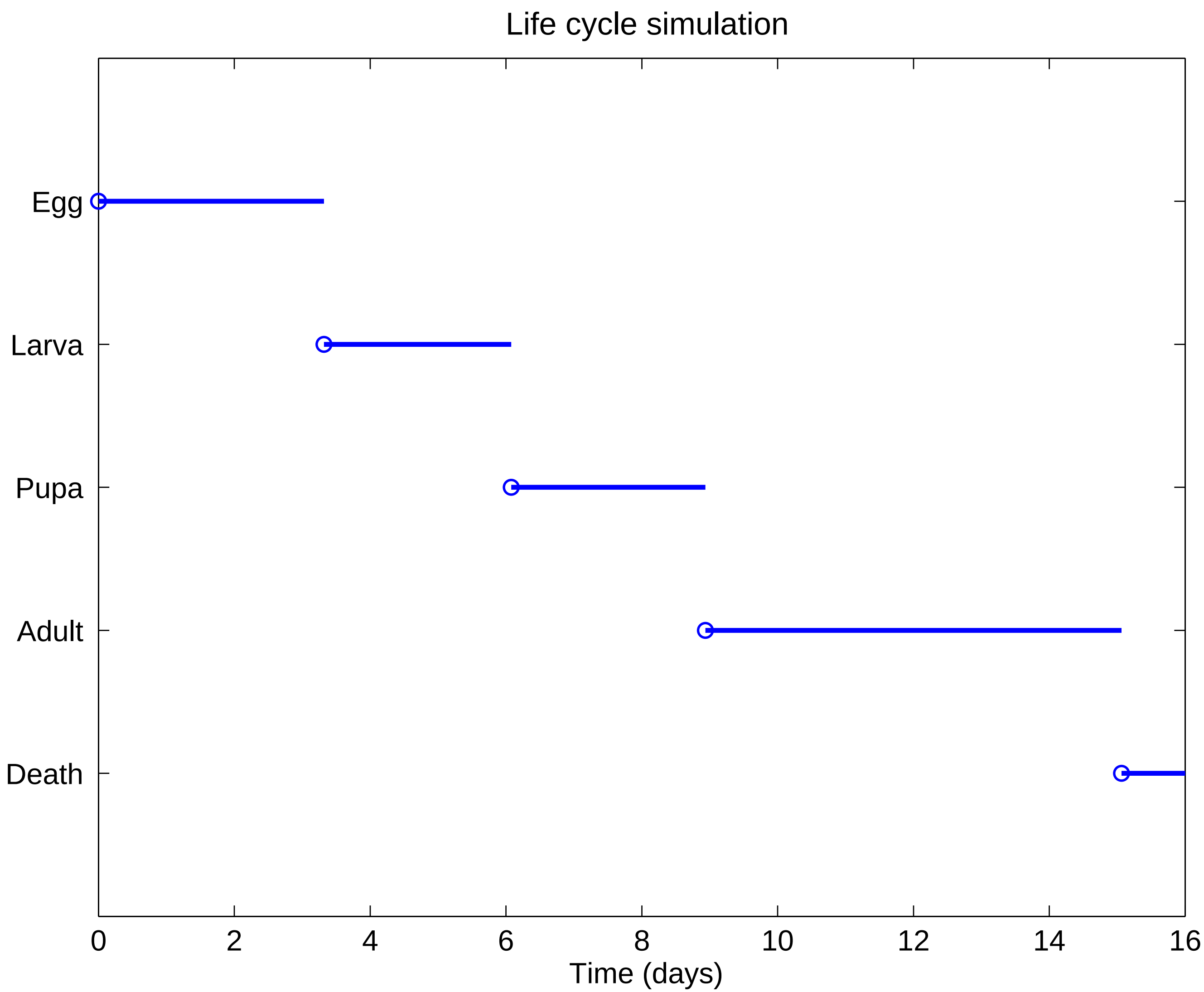
http://www.maths.uq.edu.au/~pkp

Australian Research Council
Centre of Excellence for Mathematics
and Statistics of Complex Systems

# Ensemble vs individual behaviour

Can properties of an ensemble of individuals, be deduced from a model for the behaviour of the individual?

# Butterfly life cycle



# Butterfly life cycle

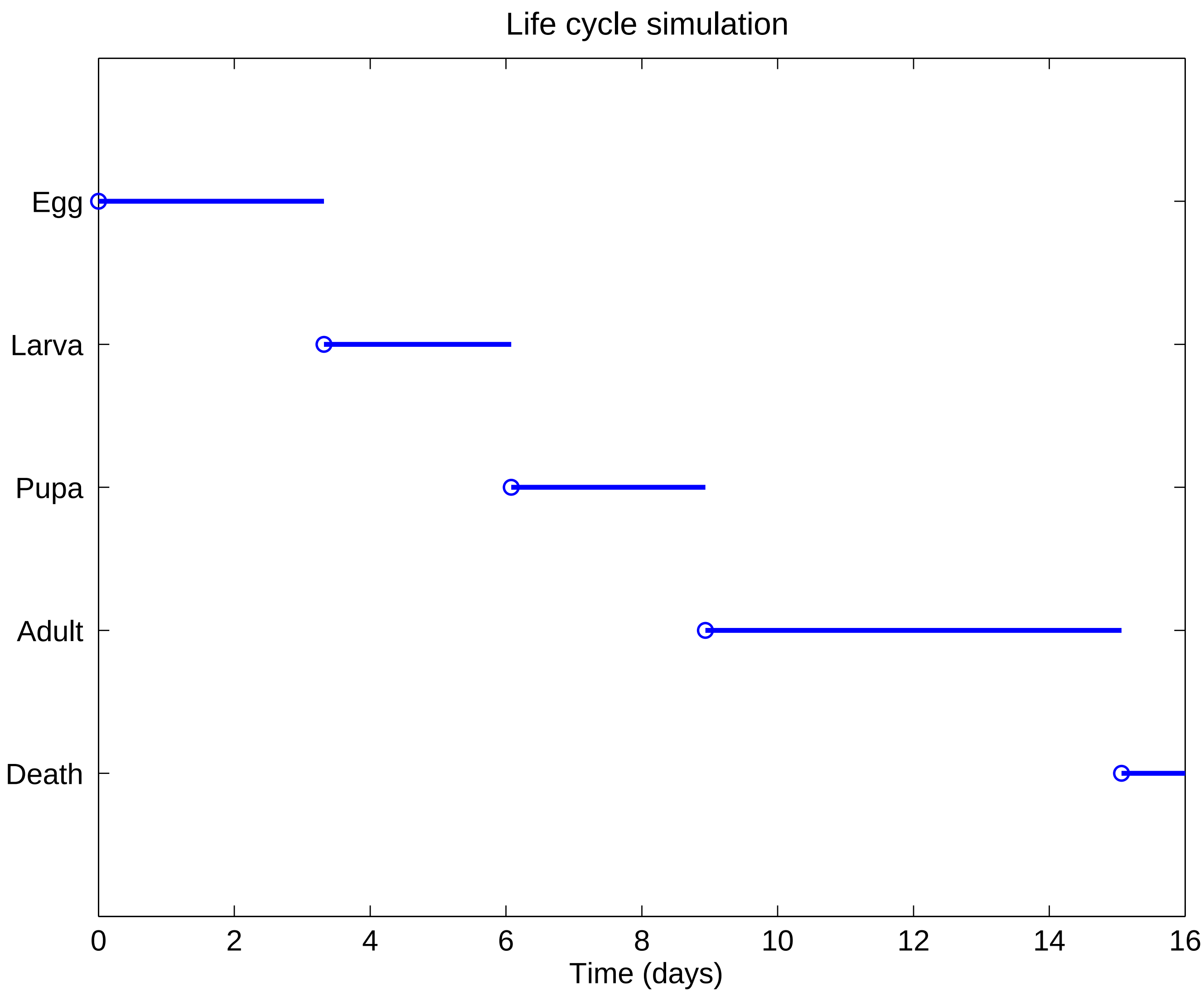
Egg $\simeq$ 4 days

Larva (caterpillar) $\simeq$ 14 days

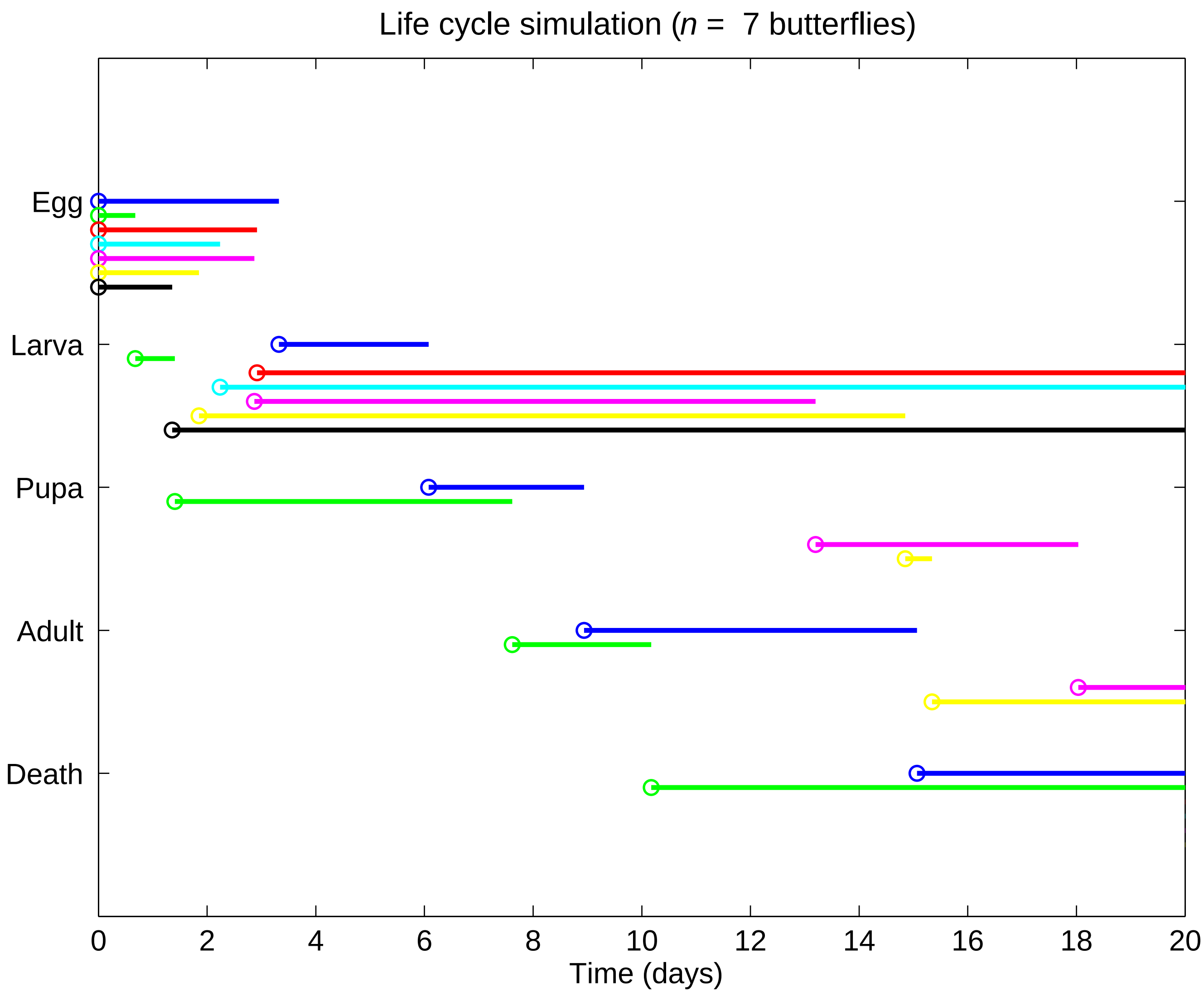
Pupa (chrysalis) $\simeq$ 7 days

Adult (butterfly) $\simeq$ 14 days

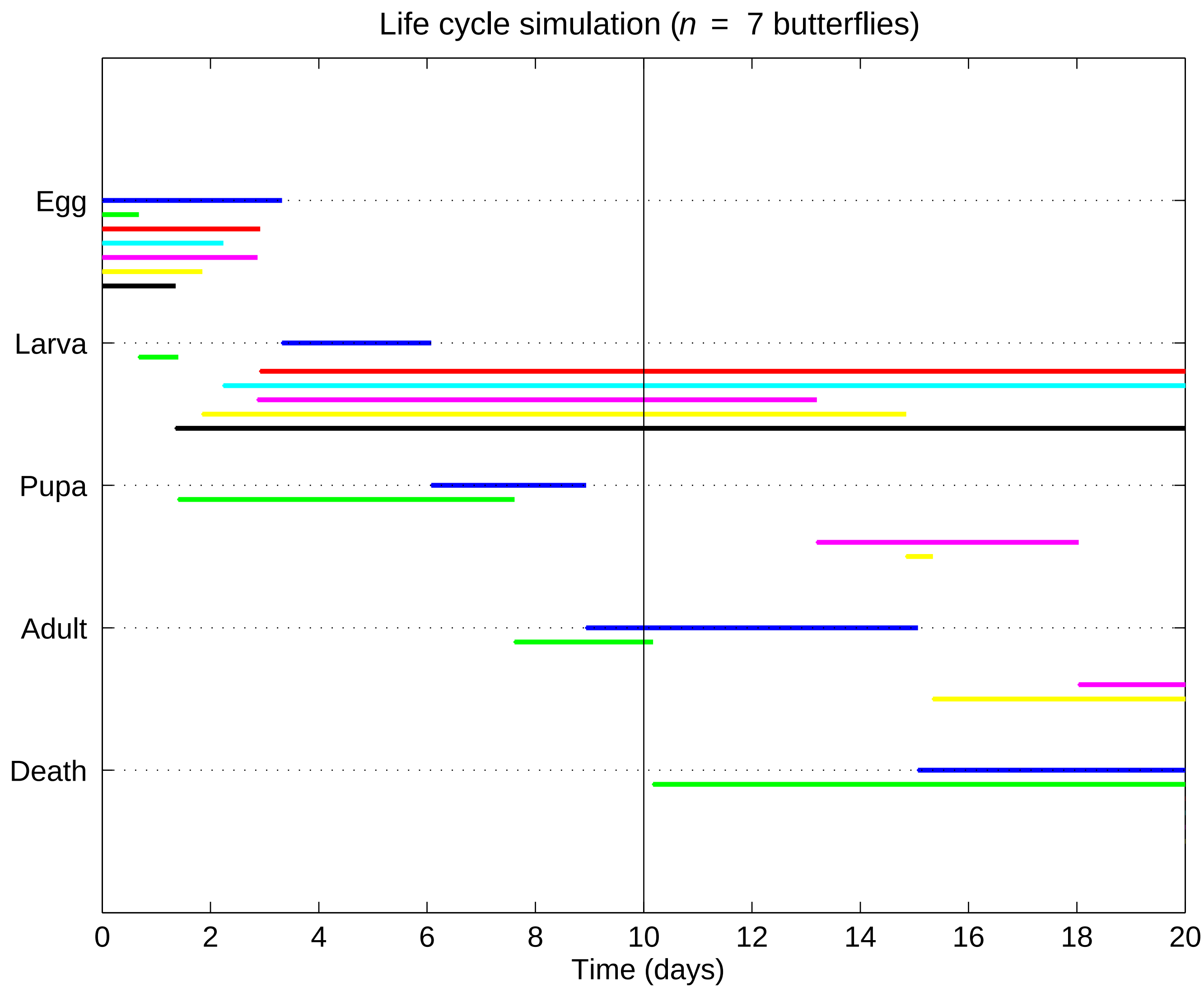
# Butterfly life cycle



# Ensemble of organisms



# Ensemble of organisms



# Ensemble vs individual behaviour

Can properties of the ensemble, be deduced from a model for the behaviour of an individual?

# Ensemble vs individual behaviour

Can properties of the ensemble, be deduced from a model for the behaviour of an individual?

For example, suppose we have $n$ butterflies.

Our intuition tells us that, for the ensemble, the *proportion* of organisms in stage $s$ at time $t$ should be approximately equal to $p_s(t)$, the *probability* that the *individual* organism is in stage $s$ at time $t$.

# Ensemble vs individual behaviour

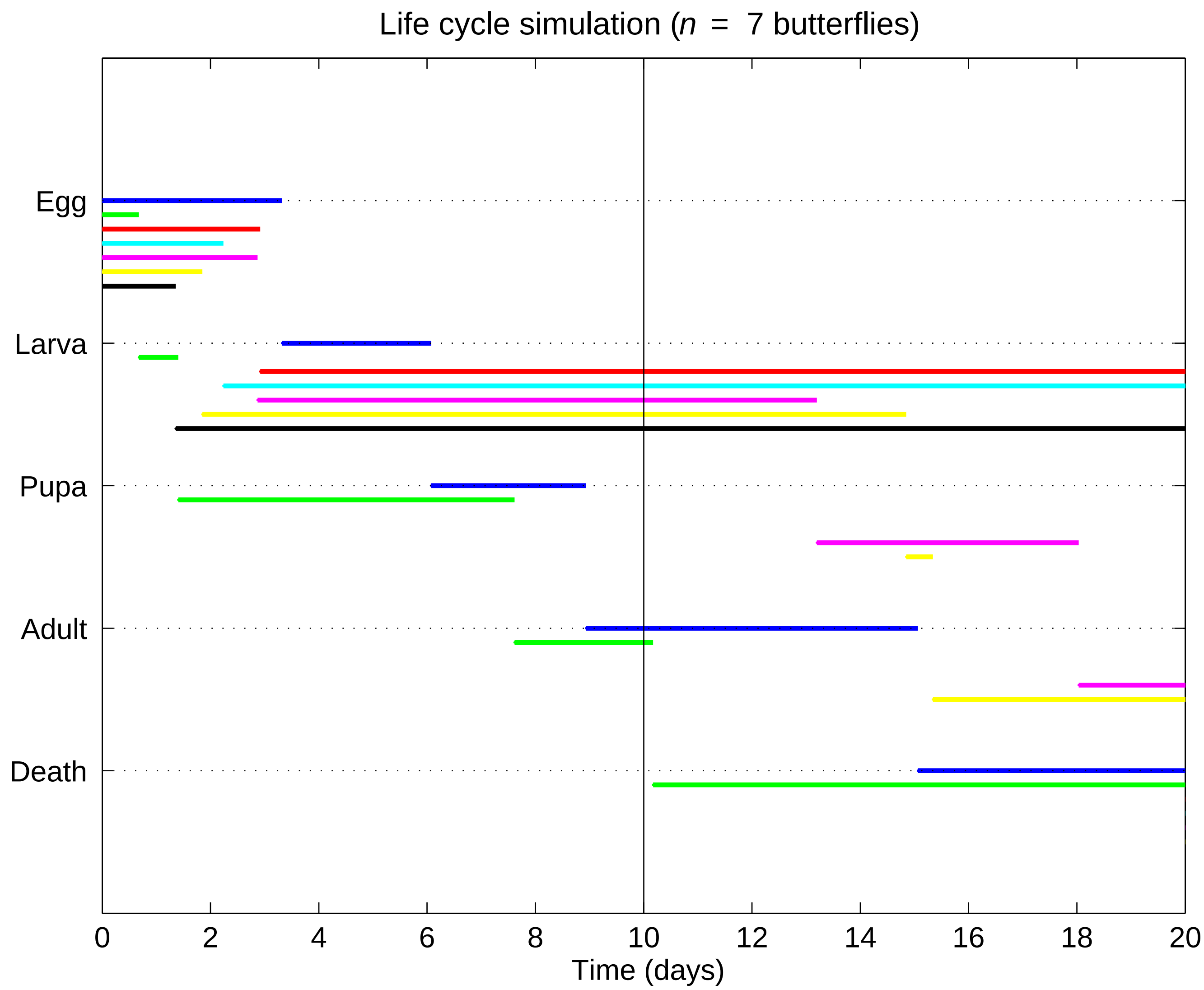
Can properties of the ensemble, be deduced from a model for the behaviour of an individual?

For example, suppose we have $n$ butterflies.

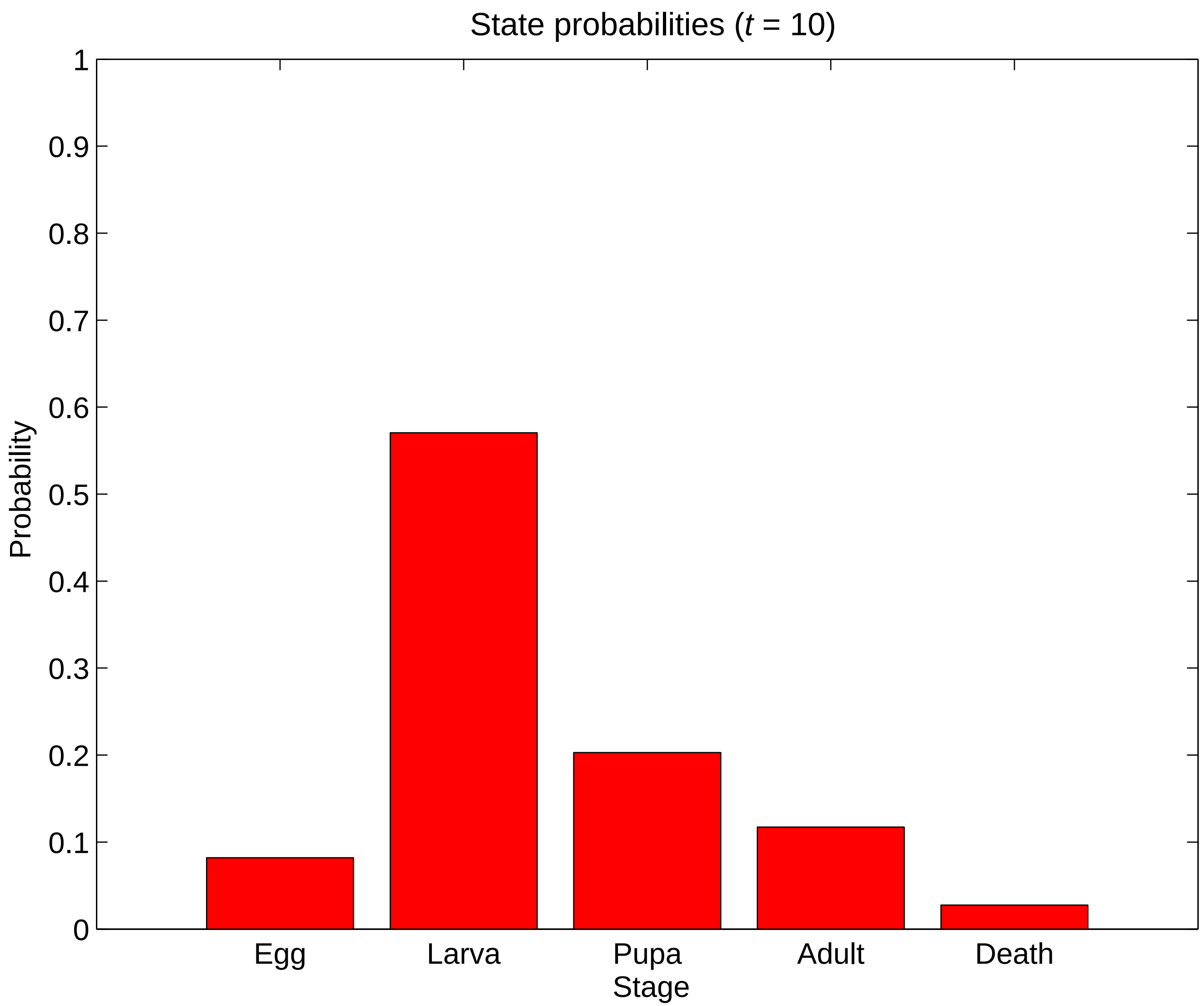
Our intuition tells us that, for the ensemble, the *proportion* of organisms in stage $s$ at time $t$ should be approximately equal to $p_s(t)$, the *probability* that the *individual* organism is in stage $s$ at time $t$.

*So strong is this intuition that scientists frequently model population proportions using individual-level models.*

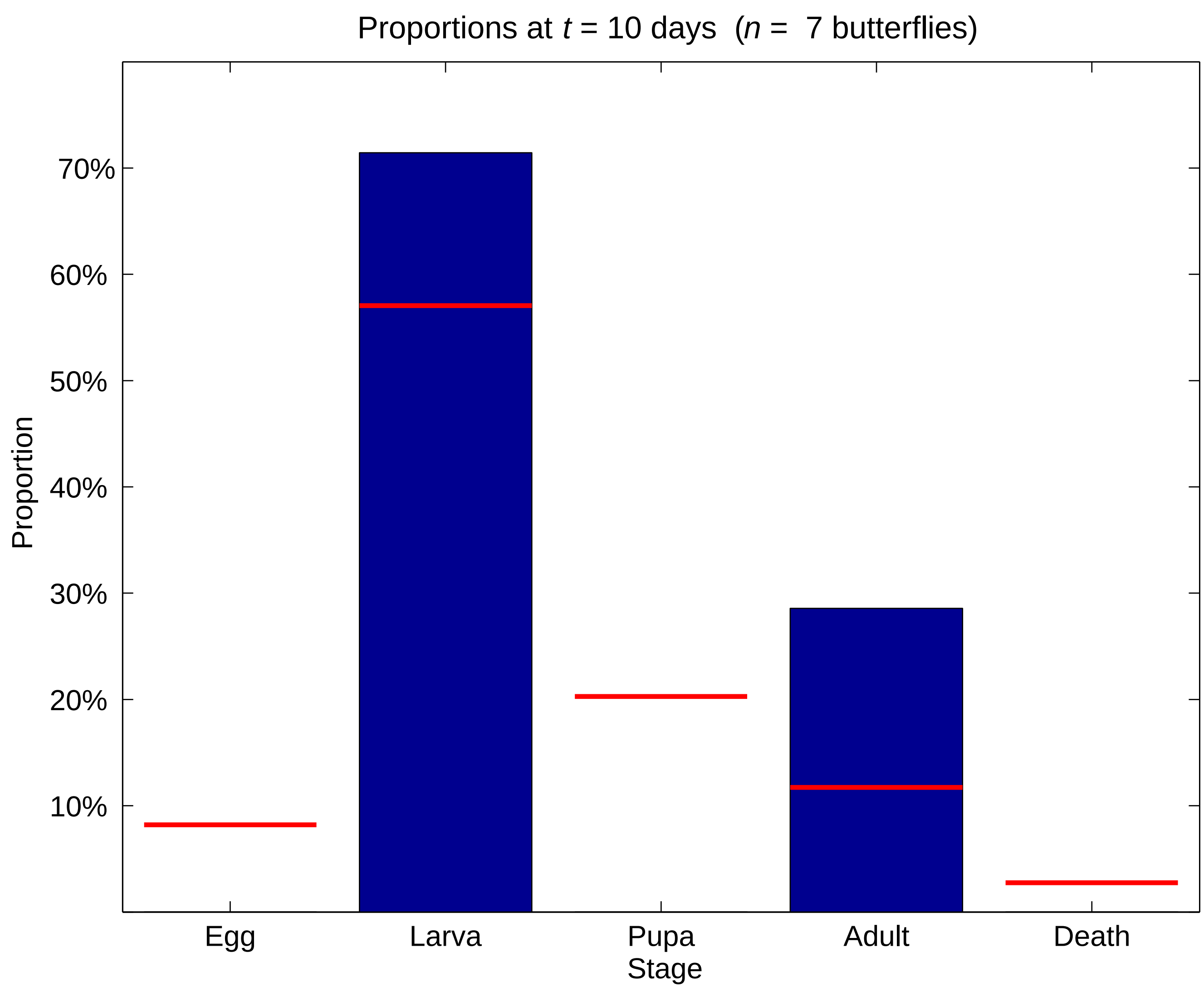
# State probabilities (individual)



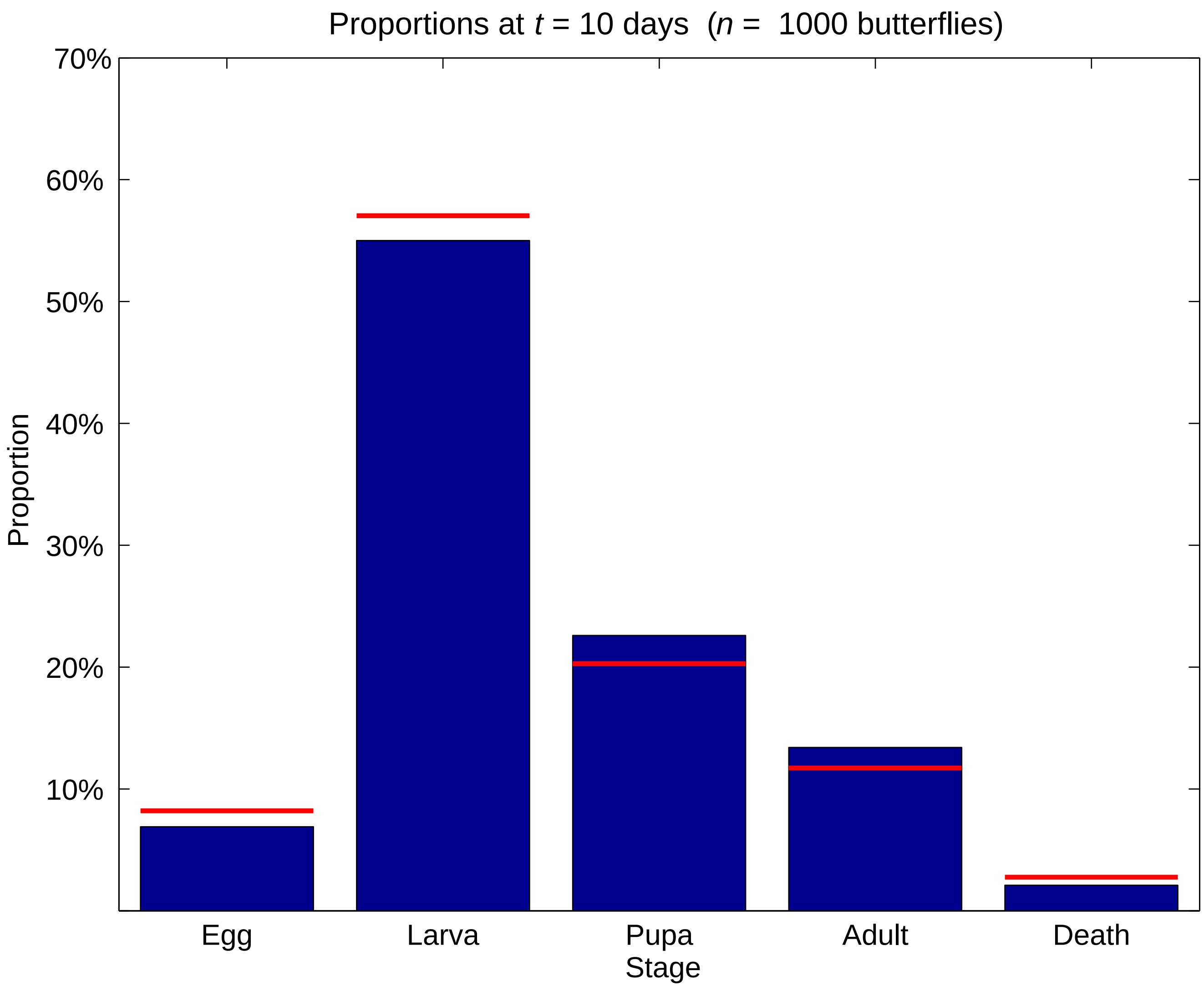
# State probabilities (individual)



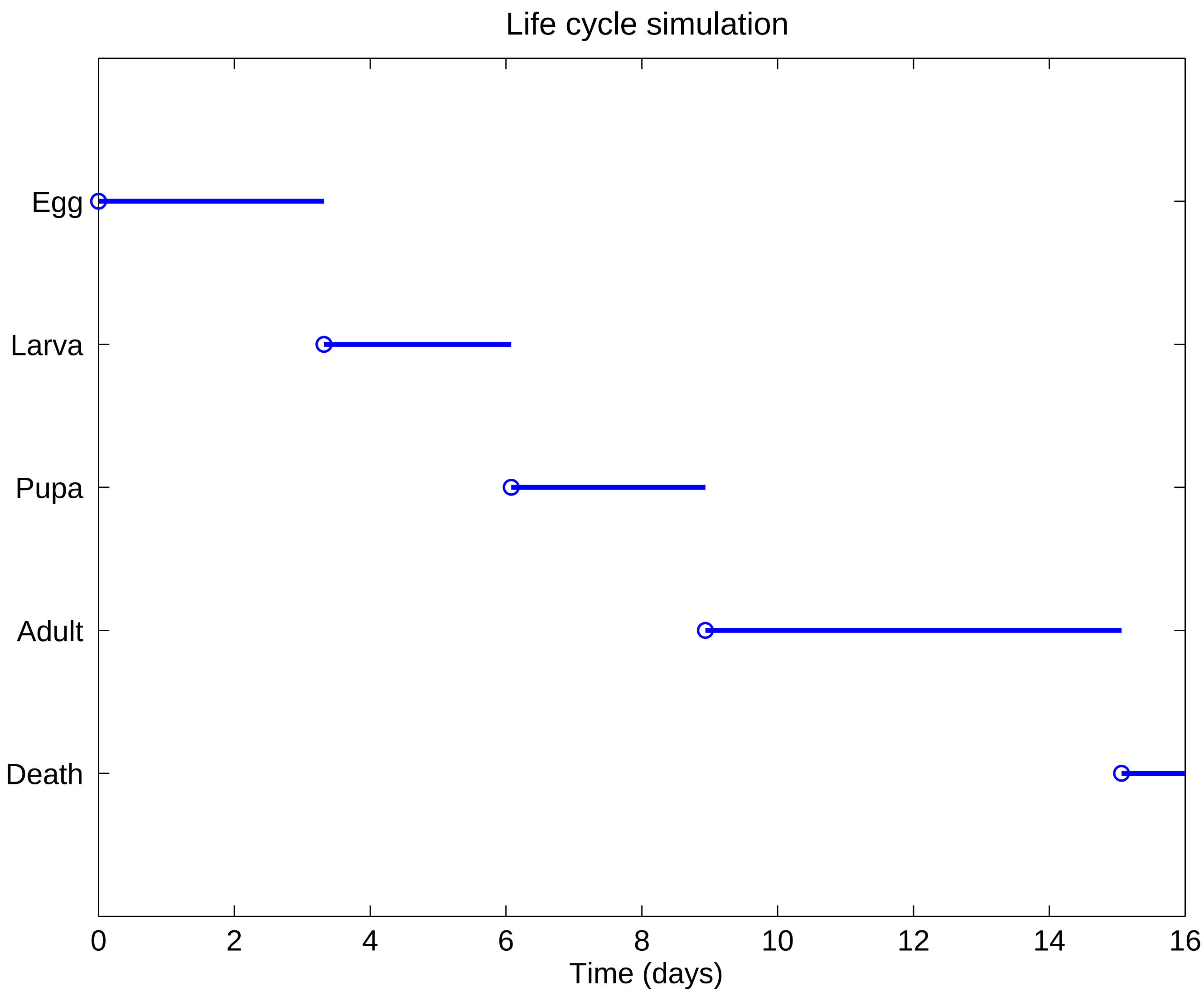
# State proportions (ensemble)



# State proportions (ensemble)



# Individual organism



# Evaluating state probabilities

What is the probability that the organism is in stage $s$ of its life cycle at time $t$?

# Evaluating state probabilities

What is the probability that the organism is in stage $s$ of its life cycle at time $t$?

Using a simple Markov chain model, we can evaluate this for each stage $s$ and for all times $t$.

# Evaluating state probabilities

$X(t)$ - the state of an individual at time $t$ $(\geq 0)$, for example, the current stage in the individual's life cycle.

Suppose $(X(t), t \geq 0)$ is a continuous-time Markov chain taking values in a discrete set $S$ with transition rates $(q_{ij})$:

- $q_{ij}$ is the rate of transition from state $i \rightarrow j$ ($j \neq i$).
- $q_i \ (= -q_{ii}) \ = \sum_{j \neq i} q_{ij}$ is the total rate out of state $i$.

# Evaluating state probabilities

$X(t)$ - the state of an individual at time $t$ $(\geq 0)$, for example, the current stage in the individual's life cycle.

Suppose $(X(t), t \geq 0)$ is a continuous-time Markov chain taking values in a discrete set $S$ with transition rates $(q_{ij})$:

$q_{ij}$ is the rate of transition from state $i \to j$ ($j \neq i$).

$q_i \ (= -q_{ii}) \ = \sum_{j \neq i} q_{ij}$ is the total rate out of state $i$.

*Example* (Butterfly life cycle) $\{4\} \to \{3\} \to \{2\} \to \{1\} \to \{0\}$

| | |
|---|---|
| $q_4 = q_{43} = 1/4$ | $\downarrow$ Egg ($\simeq$ 4 days) |
| $q_3 = q_{32} = 1/14$ | $\downarrow$ Caterpillar ($\simeq$ 14 days) |
| $q_2 = q_{21} = 1/7$ | $\downarrow$ Chrysalis ($\simeq$ 7 days) |
| $q_1 = q_{10} = 1/14$ | $\downarrow$ Adult ($\simeq$ 14 days) |

# Evaluating state probabilities

In matrix form

$$Q = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1/14 & -1/14 & 0 & 0 & 0 \\ 0 & 1/7 & -1/7 & 0 & 0 \\ 0 & 0 & 1/14 & -1/14 & 0 \\ 0 & 0 & 0 & 1/4 & -1/4 \end{pmatrix}$$

# Evaluating state probabilities

In matrix form

$$Q = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1/14 & -1/14 & 0 & 0 & 0 \\ 0 & 1/7 & -1/7 & 0 & 0 \\ 0 & 0 & 1/14 & -1/14 & 0 \\ 0 & 0 & 0 & 1/4 & -1/4 \end{pmatrix}$$

Why put minus the total rate on the diagonal?

# Evaluating state probabilities

In matrix form

$$Q = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1/14 & -1/14 & 0 & 0 & 0 \\ 0 & 1/7 & -1/7 & 0 & 0 \\ 0 & 0 & 1/14 & -1/14 & 0 \\ 0 & 0 & 0 & 1/4 & -1/4 \end{pmatrix}$$

Why put minus the total rate on the diagonal?

For mathematical convenience . . . the equations we must solve are then easier to write down.

# Evaluating state probabilities

The state probabilities $\boldsymbol{p}(t) = (p_j(t), j \in S)$, where

$$p_j(t) = \Pr(X(t) = j),$$

can be obtained as the (unique) solution to

$$\boldsymbol{p}'(t) = \boldsymbol{p}(t)\,Q \quad \text{satisfying} \quad \boldsymbol{p}(0) = \boldsymbol{a},$$

where $\boldsymbol{a} = (a_j, j \in S)$ is a given initial distribution.

# Evaluating state probabilities

The state probabilities $\boldsymbol{p}(t) = (p_j(t), j \in S)$, where

$$p_j(t) = \Pr(X(t) = j),$$

can be obtained as the (unique) solution to

$$\boldsymbol{p}'(t) = \boldsymbol{p}(t)\, Q \quad \text{satisfying} \quad \boldsymbol{p}(0) = \boldsymbol{a},$$

where $\boldsymbol{a} = (a_j, j \in S)$ is a given initial distribution.

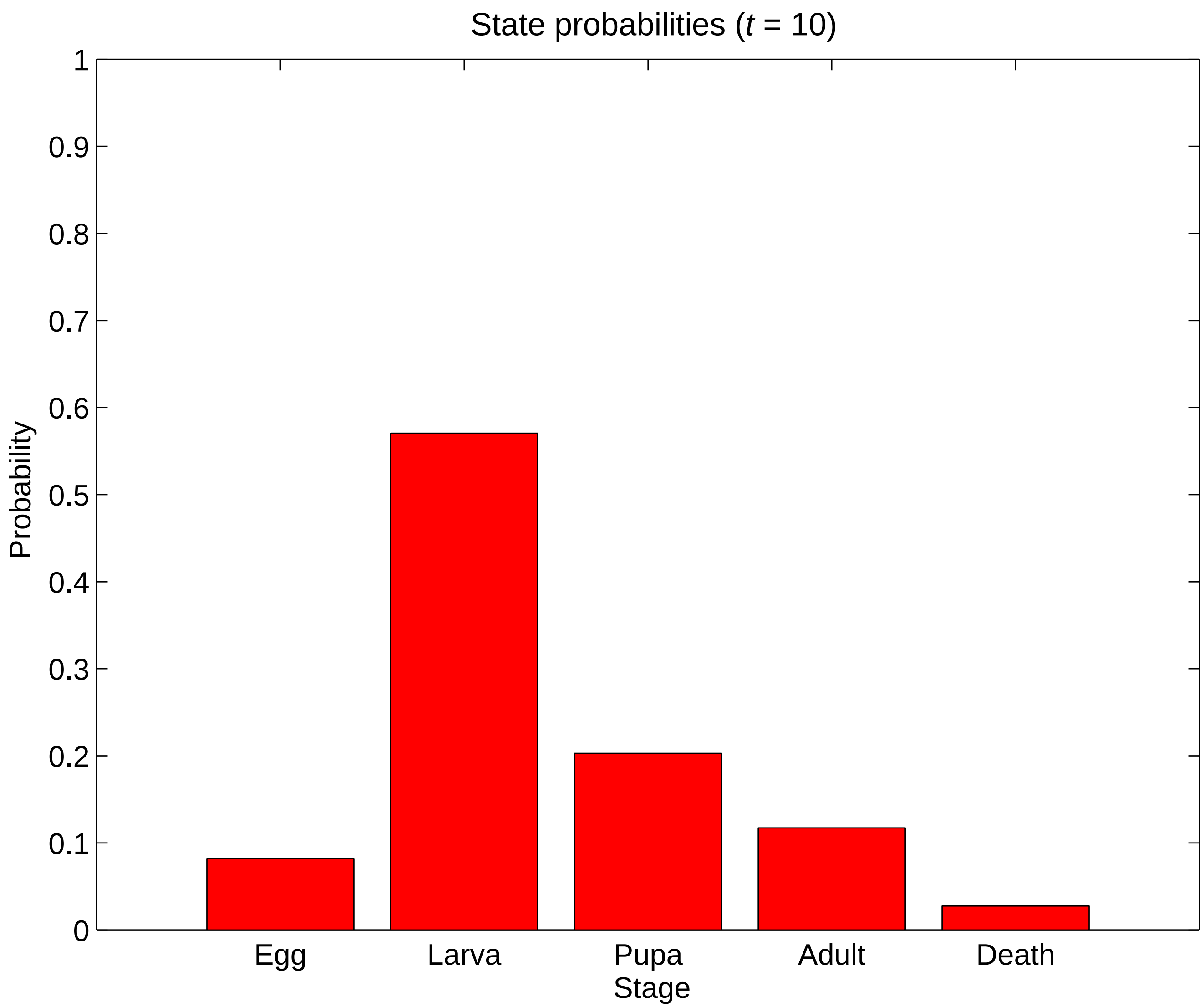
(It will be convenient to restrict our attention to the case where $S$ is a *finite* set, but I note that many of the arguments presented hold more generally.)

# Using Matlab

```
% State probabilities (butterfly life cycle)

q(1)=1/14; q(2)=1/7; q(3)=1/14; q(4)=1/4;
Q=zeros(5,5);
for i=2:5
  state=i-1; % Matlab doesn't like a 0 index
  Q(i,i-1)=q(state); Q(i,i)=-q(state);
end
i=5; t=10;
P=expm(Q*t); % The solution to p'(t)=p(t)Q
p=P(i,1:5);  % with p_4(0)=1
bar(0:4,p);
```

# Individual organism



# Analytically

Sometimes the state probabilities can be evaluated analytically.

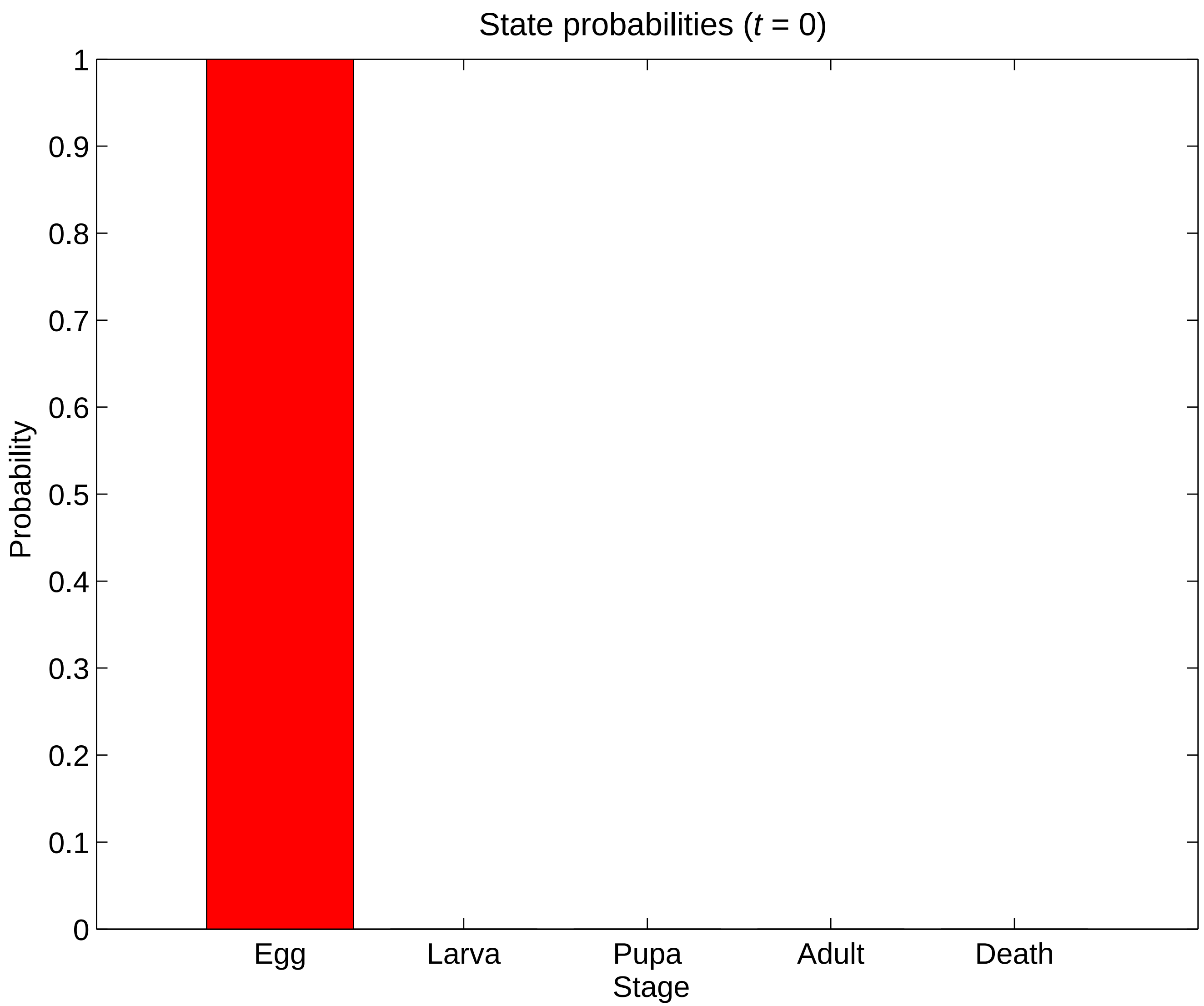
Suppose that an organism has $M$ stages of life ($M = 4$ for the butterfly), and that the expected time spent in stage $j$ is $1/q_j$ ($q_j$ is the rate of departure from stage $j$).

If $q_1, q_2, \ldots, q_M$ are distinct, then

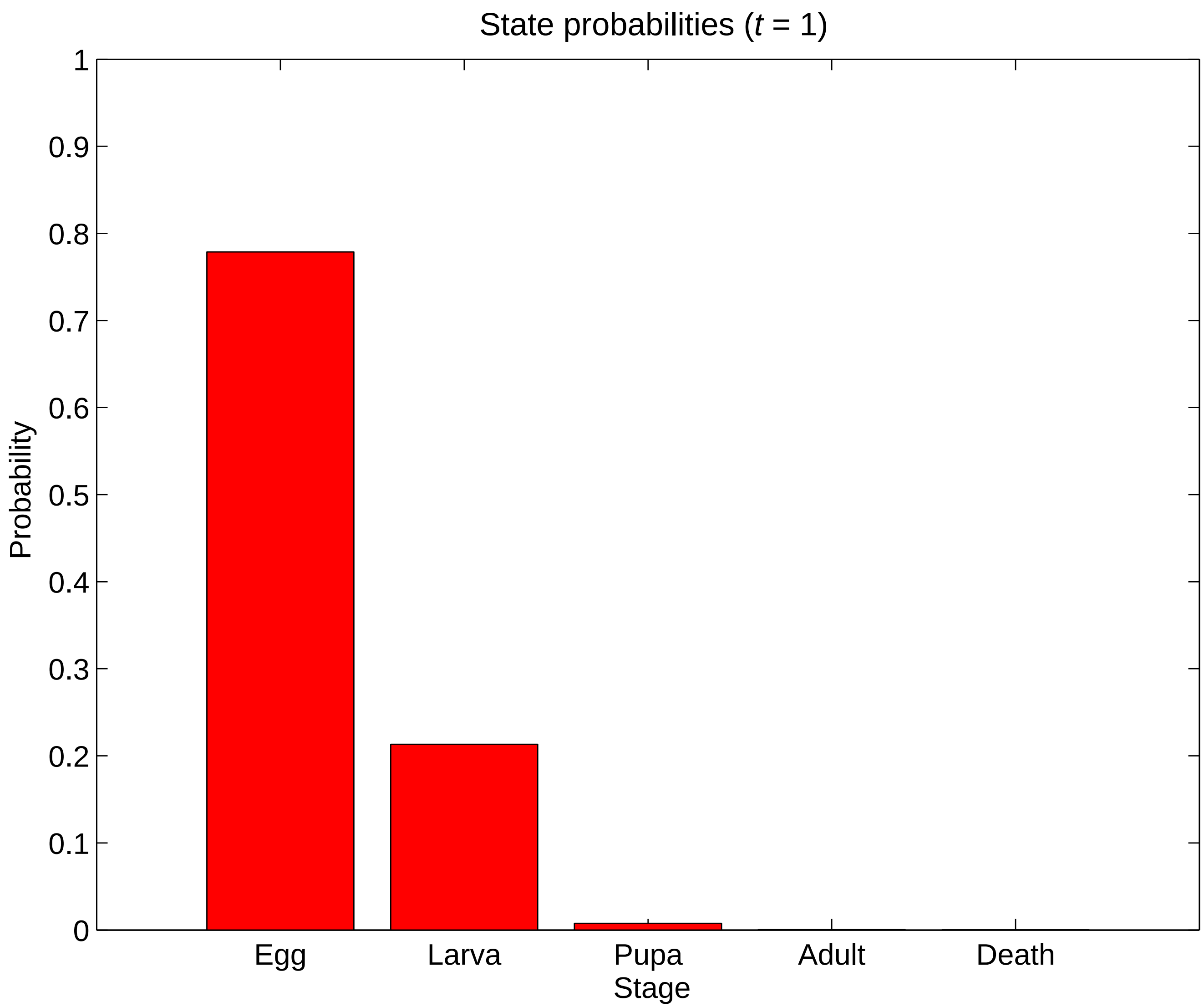
$$p_j(t) = \frac{1}{q_j} \sum_{k=j}^{M} q_k e^{-q_k t} \prod_{l=j,\, l \neq k}^{M} \frac{q_l}{q_l - q_k},$$

for $j = 1, \ldots, M$, and $p_0(t) = 1 - \sum_{j=1}^{M} p_j(t)$.

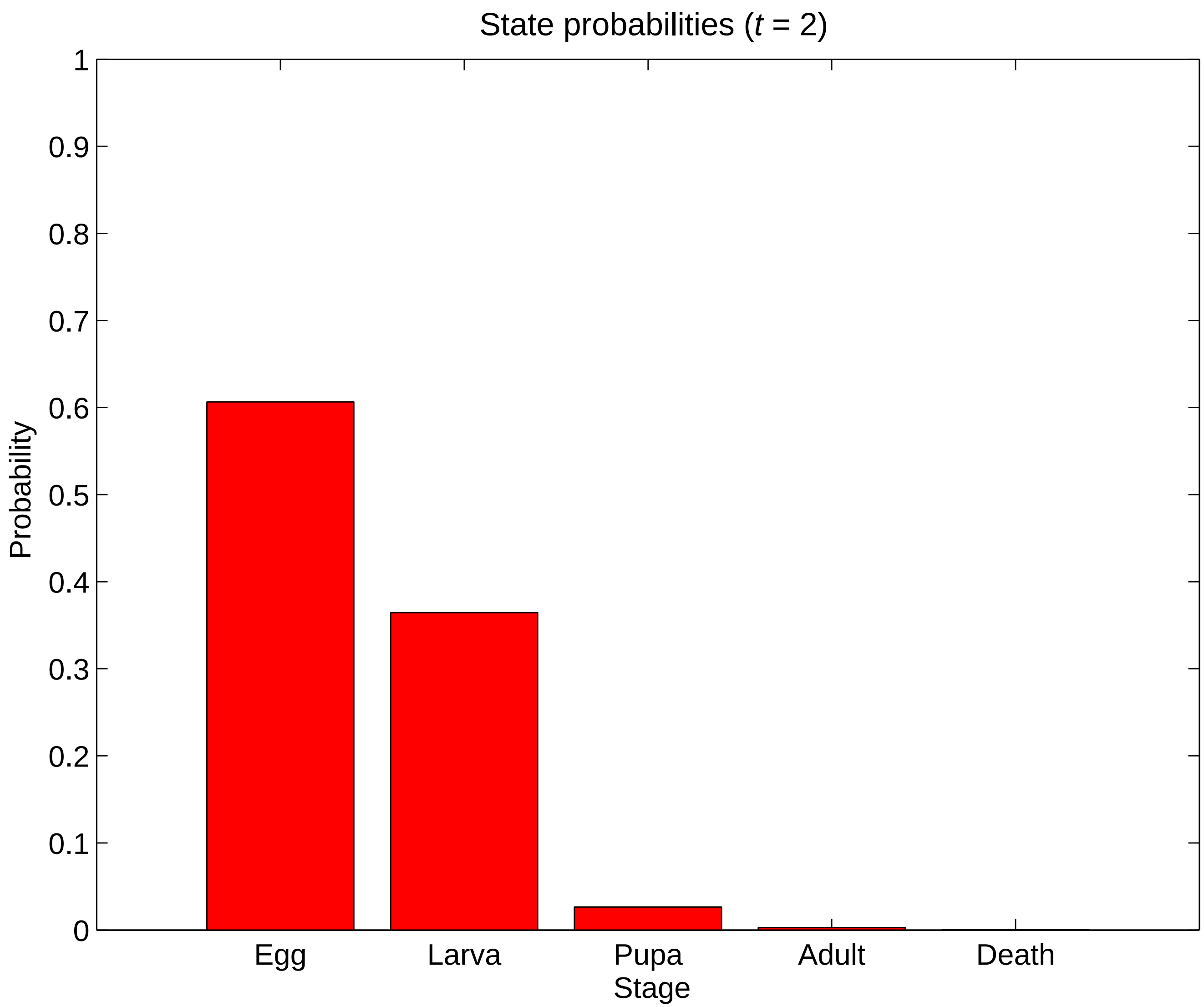
# Individual organism



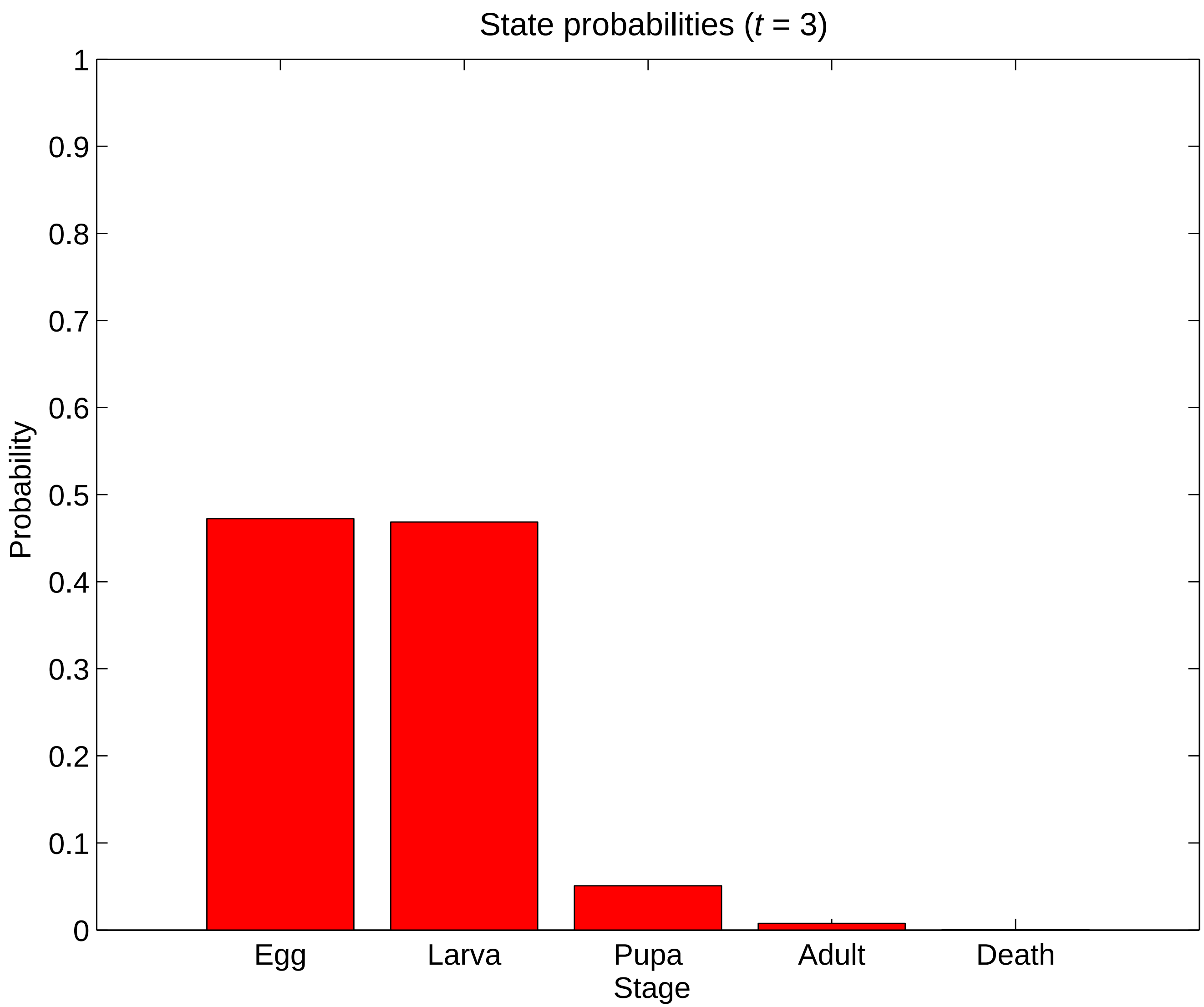
# Individual organism



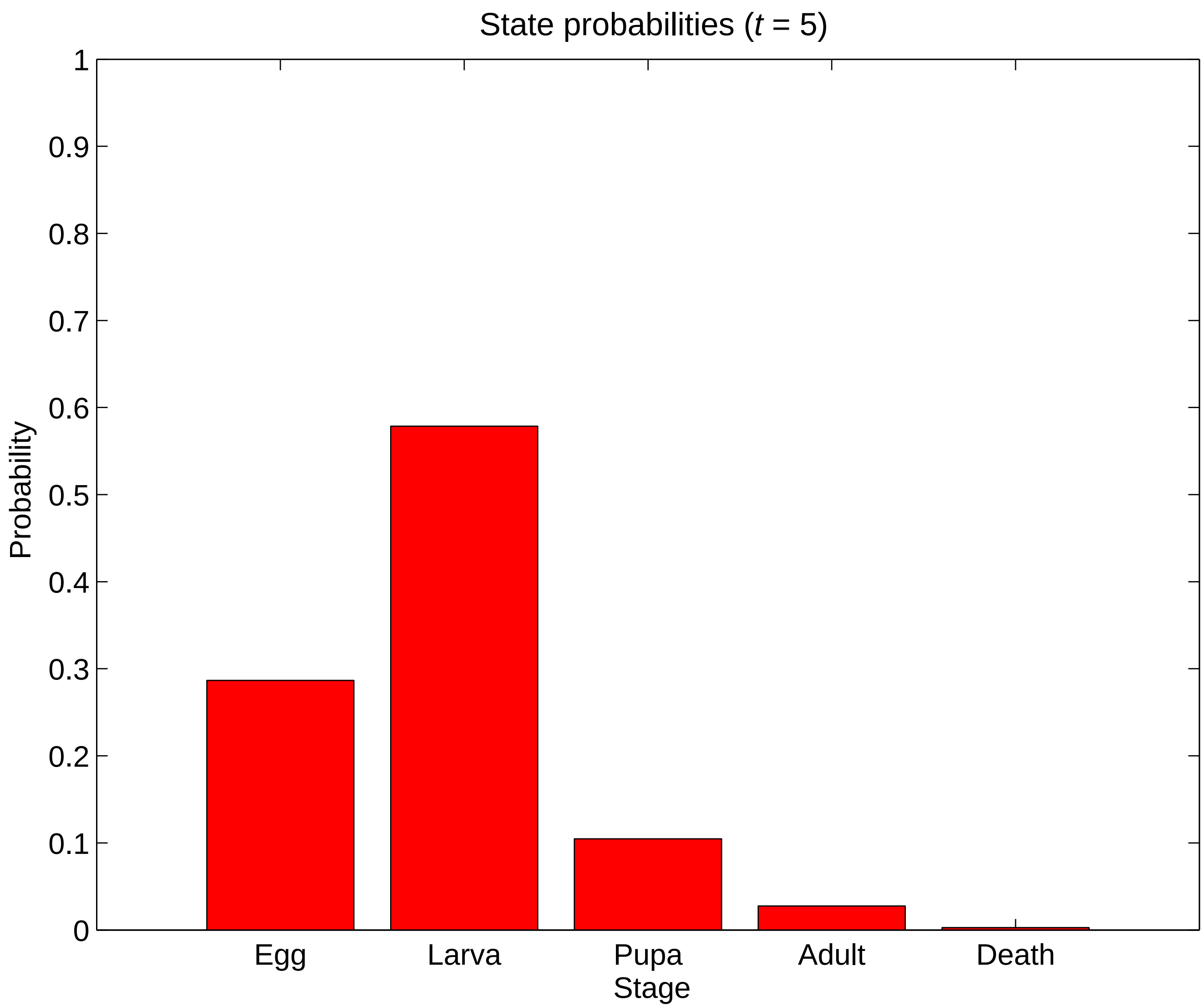
# Individual organism



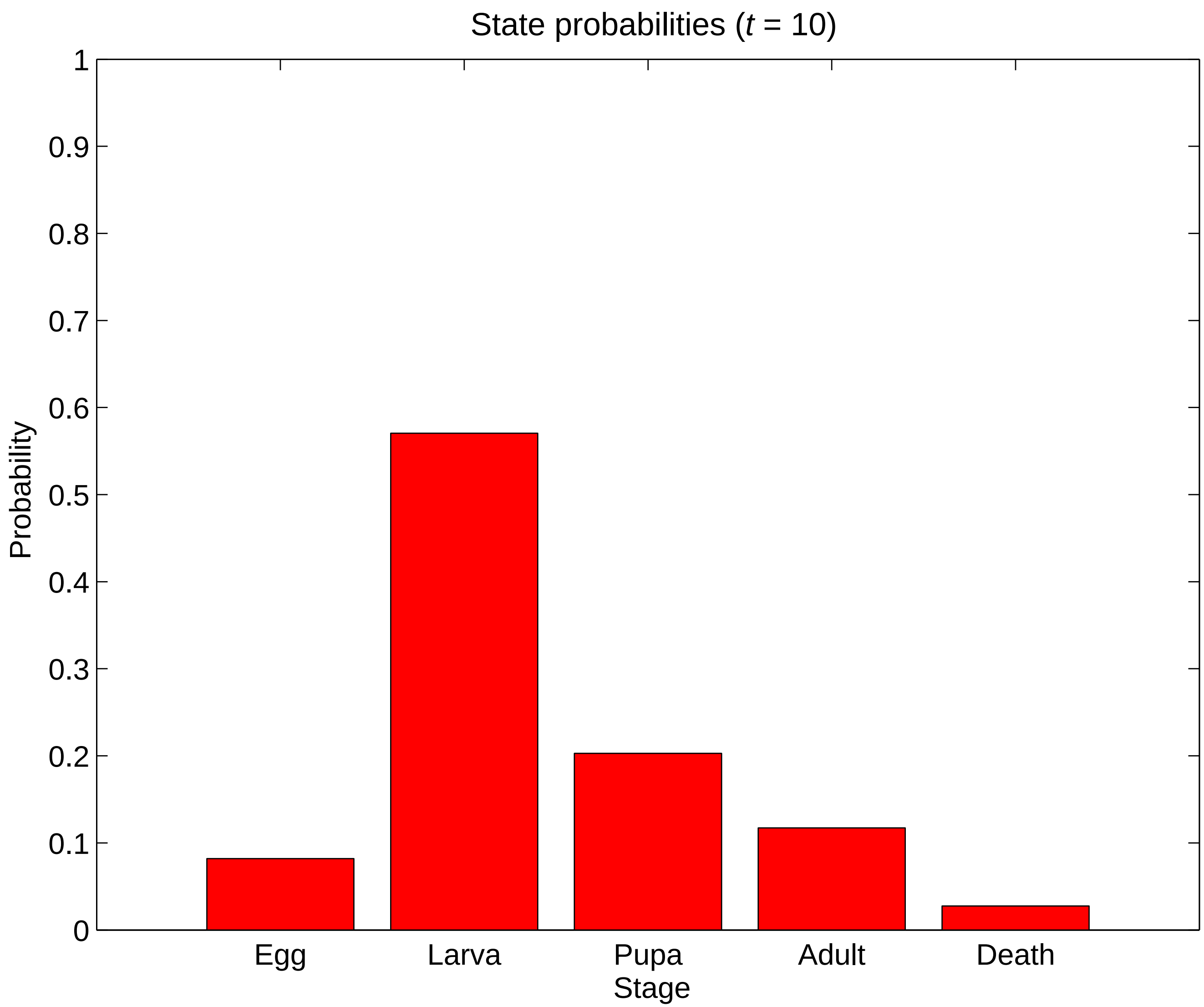
# Individual organism



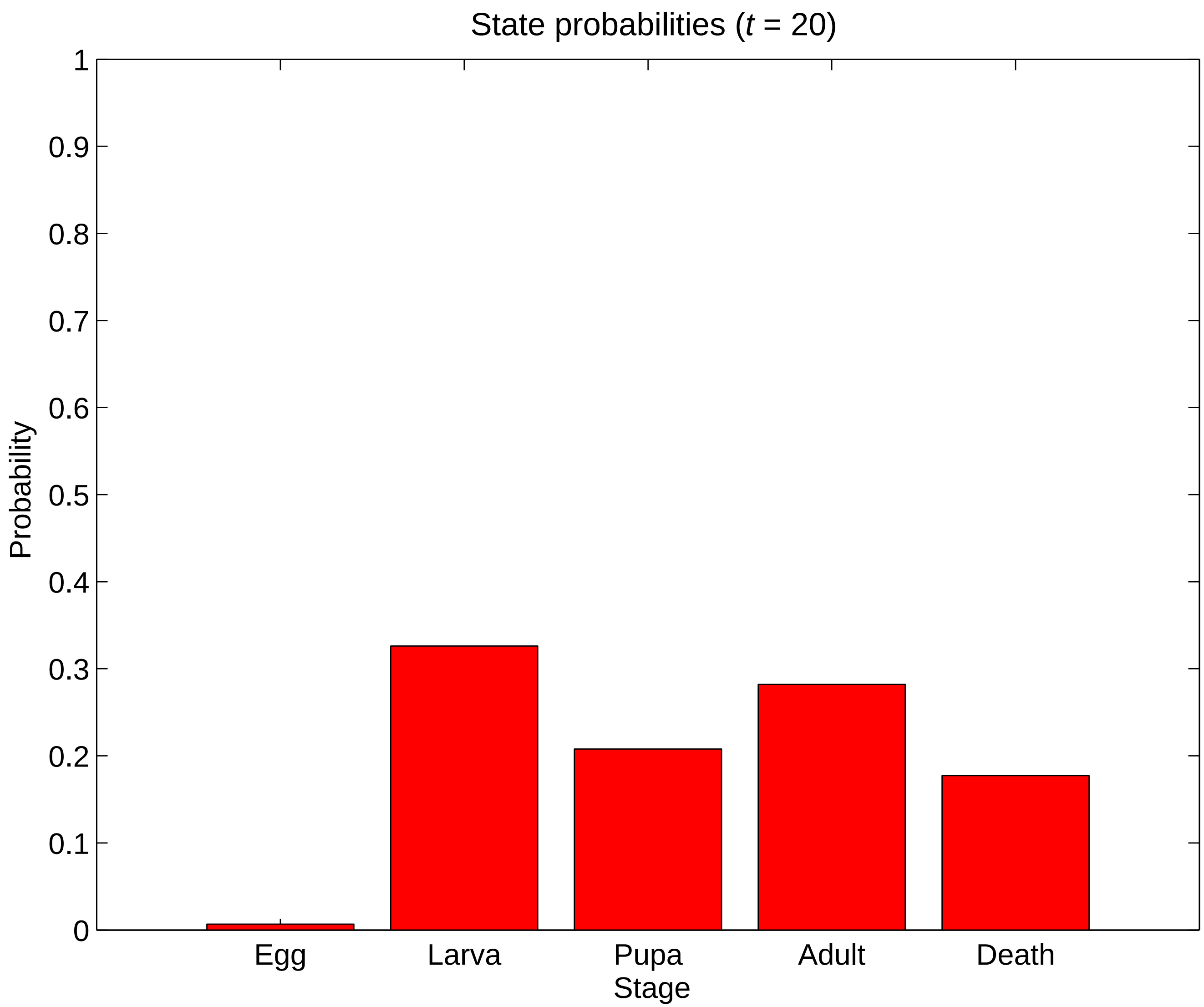
# Individual organism



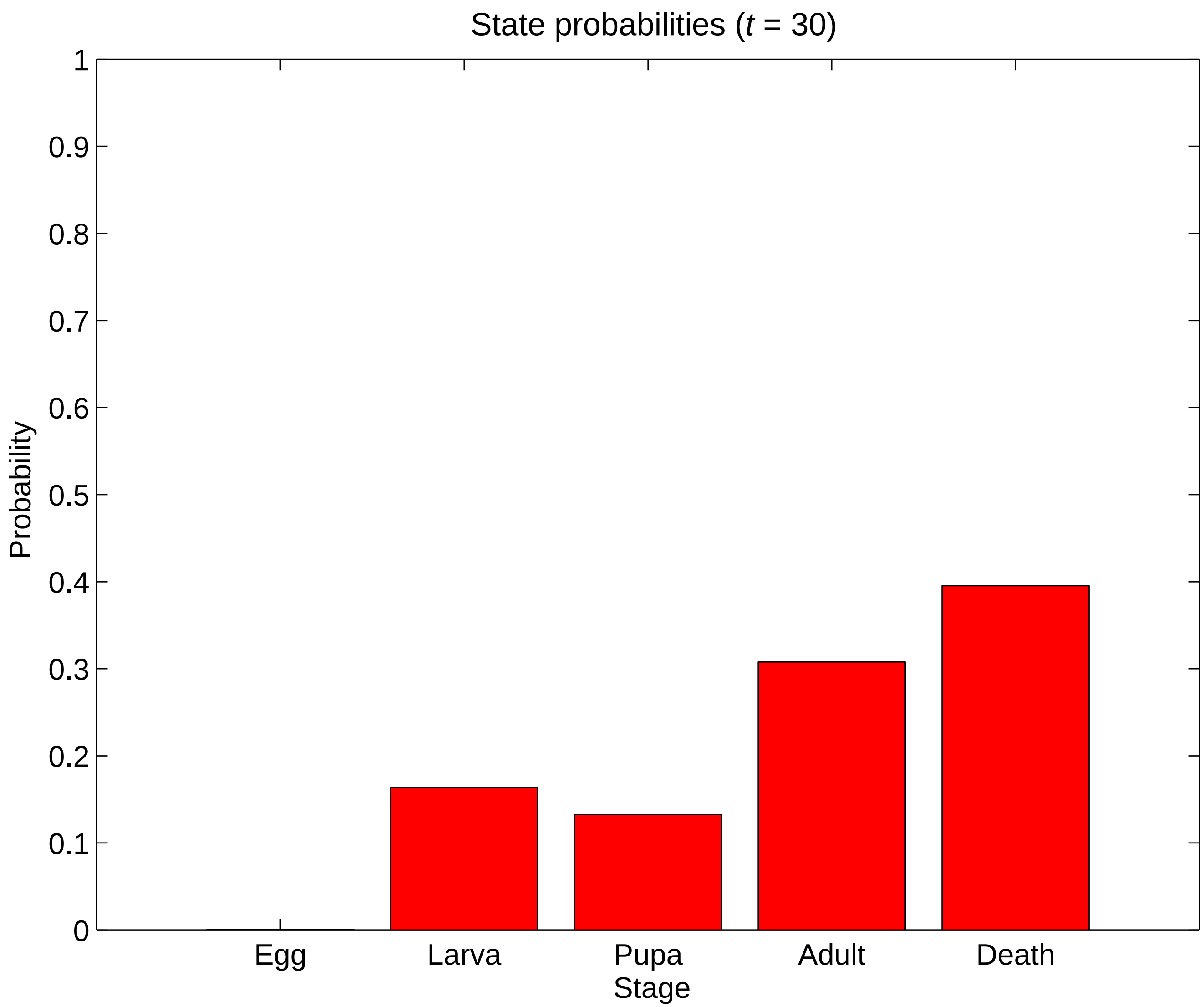
# Individual organism



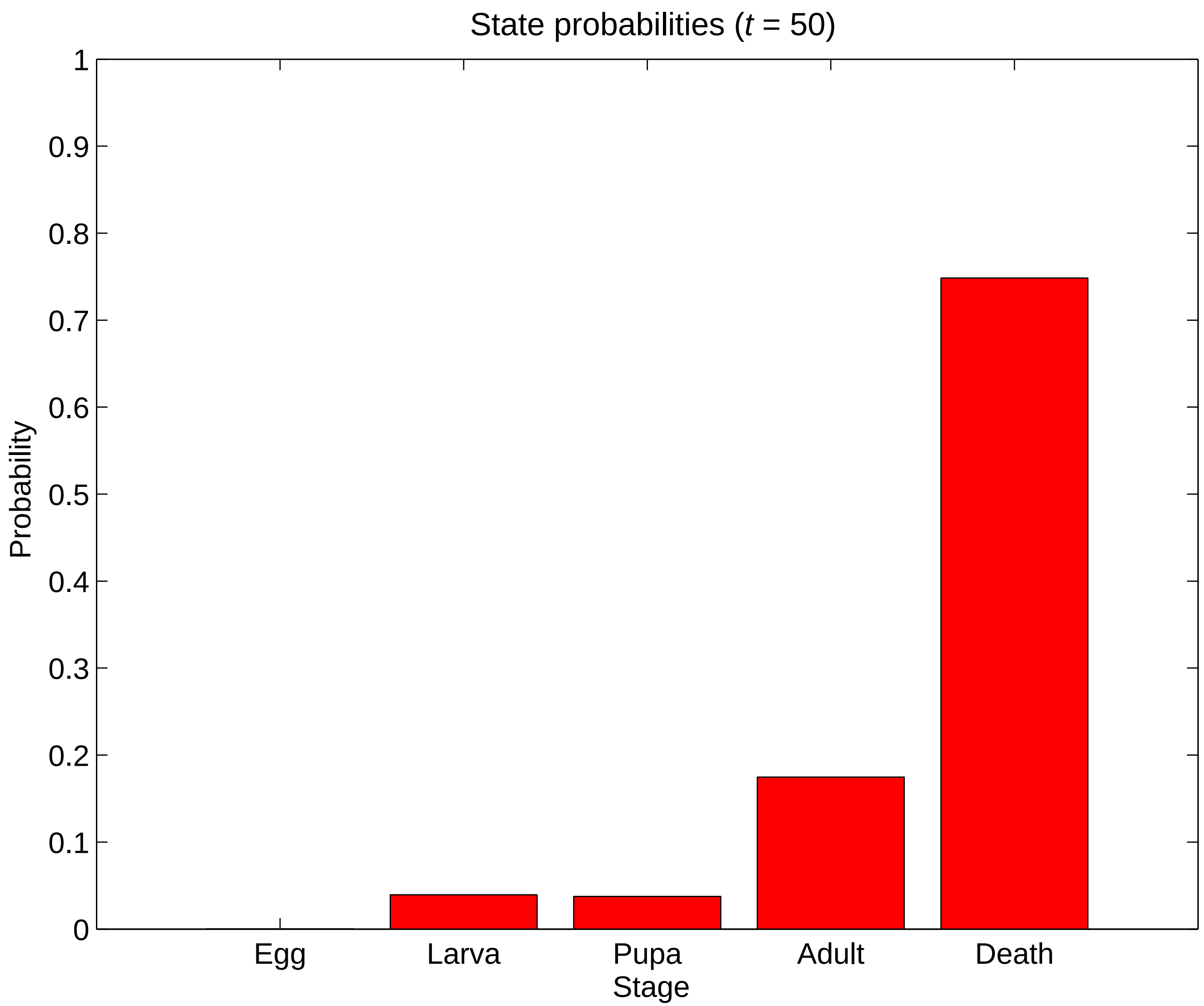
# Individual organism



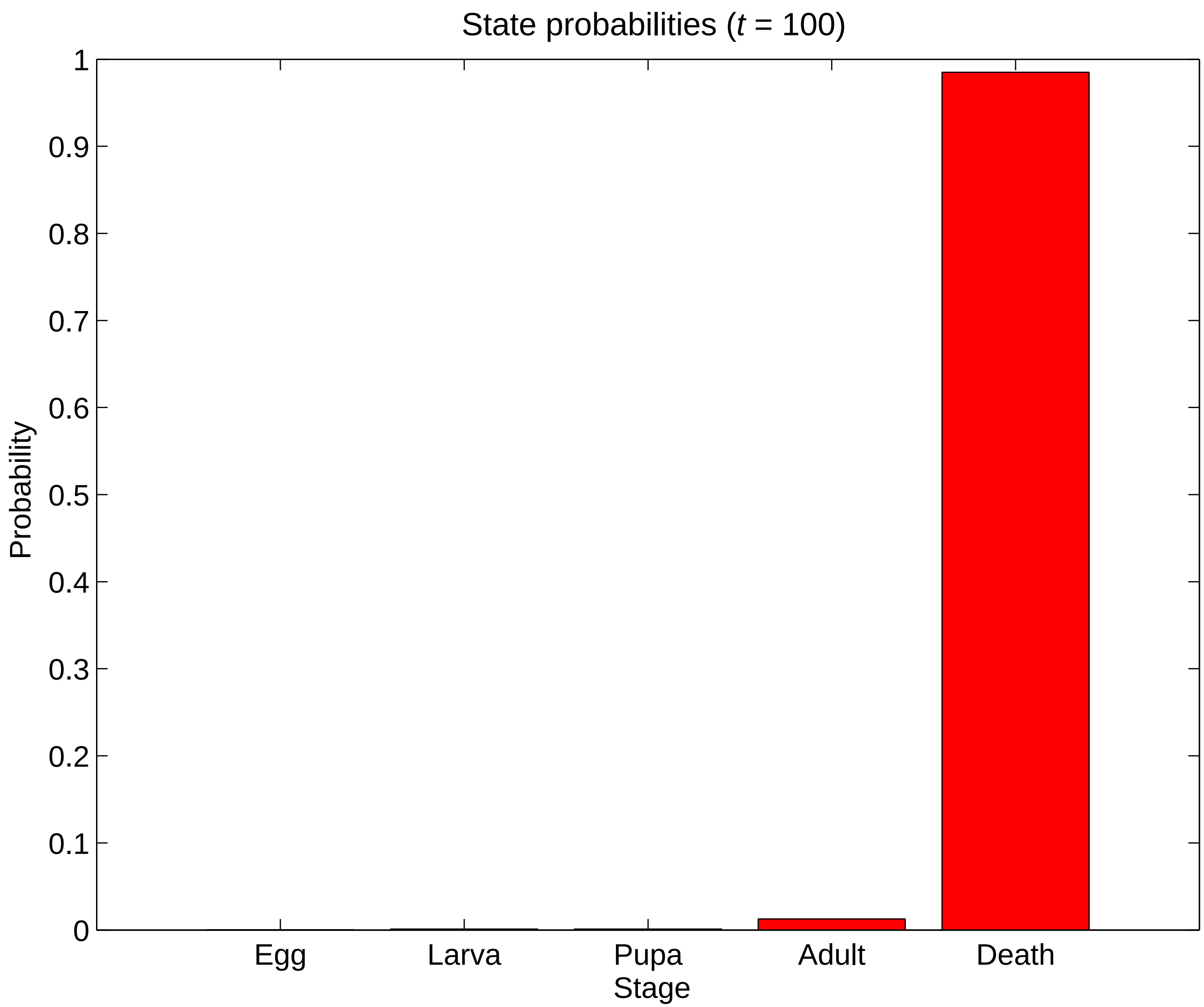
# Individual organism



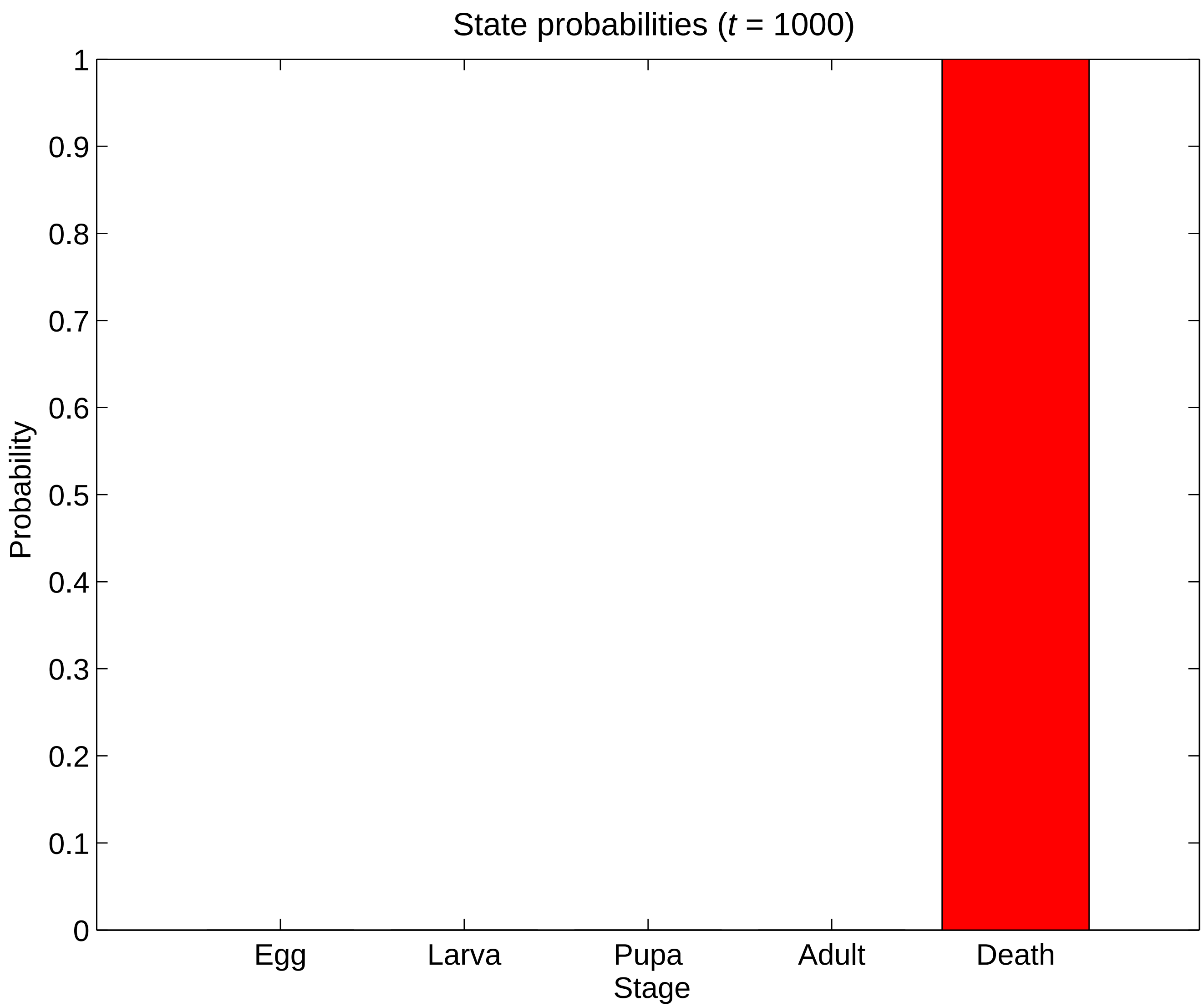
# Individual organism



# Individual organism



# Individual organism



# Ensemble of organisms

# The ensemble model

Suppose that at time $t = 0$ the individuals are assigned to the states according to some rule and then each moves independently in $S$ as a Markov chain governed by $Q$.

The key assumption here is *independence*: individuals do not affect one another.

We record only the *number* of individuals in the various states, rather than their positions.

Let $N_j(t)$ be the number of individuals in state $j$ at time $t$, and let $\boldsymbol{N} = (N_j, j \in S)$. The process $(\boldsymbol{N}(t), t \geq 0)$ is also a continuous-time Markov chain.

# The ensemble model

Suppose that at time $t = 0$ the individuals are assigned to the states according to some rule and then each moves independently in $S$ as a Markov chain governed by $Q$.

The key assumption here is *independence*: individuals do not affect one another.

We record only the *number* of individuals in the various states, rather than their positions.

Let $N_j(t)$ be the number of individuals in state $j$ at time $t$, and let $\boldsymbol{N} = (N_j, j \in S)$. The process $(\boldsymbol{N}(t), t \geq 0)$ is also a continuous-time Markov chain.

# The ensemble model

Suppose that at time $t = 0$ the individuals are assigned to the states according to some rule and then each moves independently in $S$ as a Markov chain governed by $Q$.

The key assumption here is *independence*: individuals do not affect one another.

We record only the *number* of individuals in the various states, rather than their positions.

Let $N_j(t)$ be the number of individuals in state $j$ at time $t$, and let $\boldsymbol{N} = (N_j, j \in S)$. The process $(\boldsymbol{N}(t), t \geq 0)$ is also a continuous-time Markov chain.

# The closed ensemble

*The closed ensemble*. We suppose that there is a *fixed* number $n$ of individuals, each moving according to $Q$.

The process takes values in

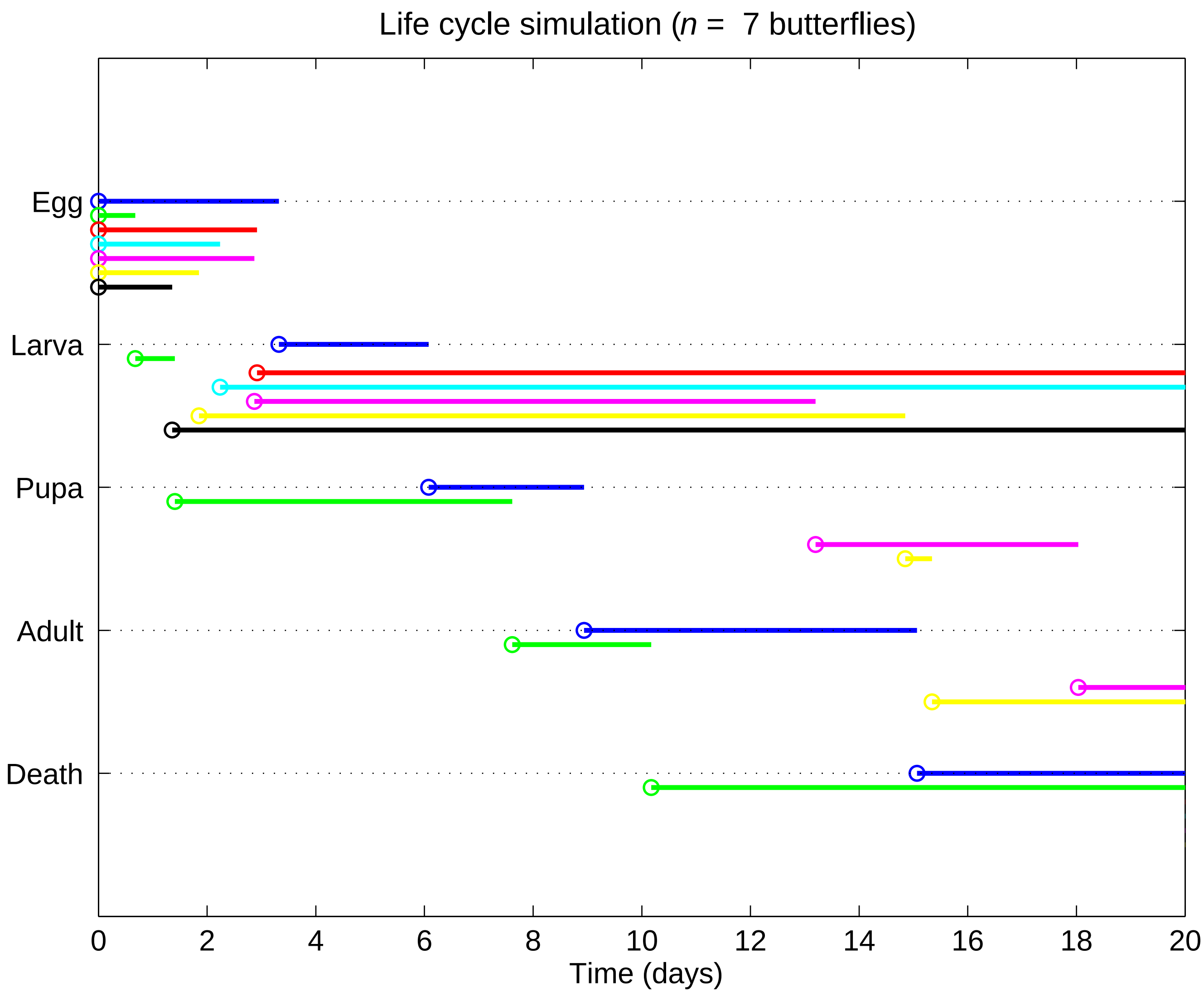
$$E = \{\boldsymbol{n} \in \{0, \ldots, n\}^S : \textstyle\sum_{j \in S} n_j = n\},$$

and its transition rates $Q_{E} = (q(\boldsymbol{n}, \boldsymbol{m}), \boldsymbol{n}, \boldsymbol{m} \in E)$ are given by

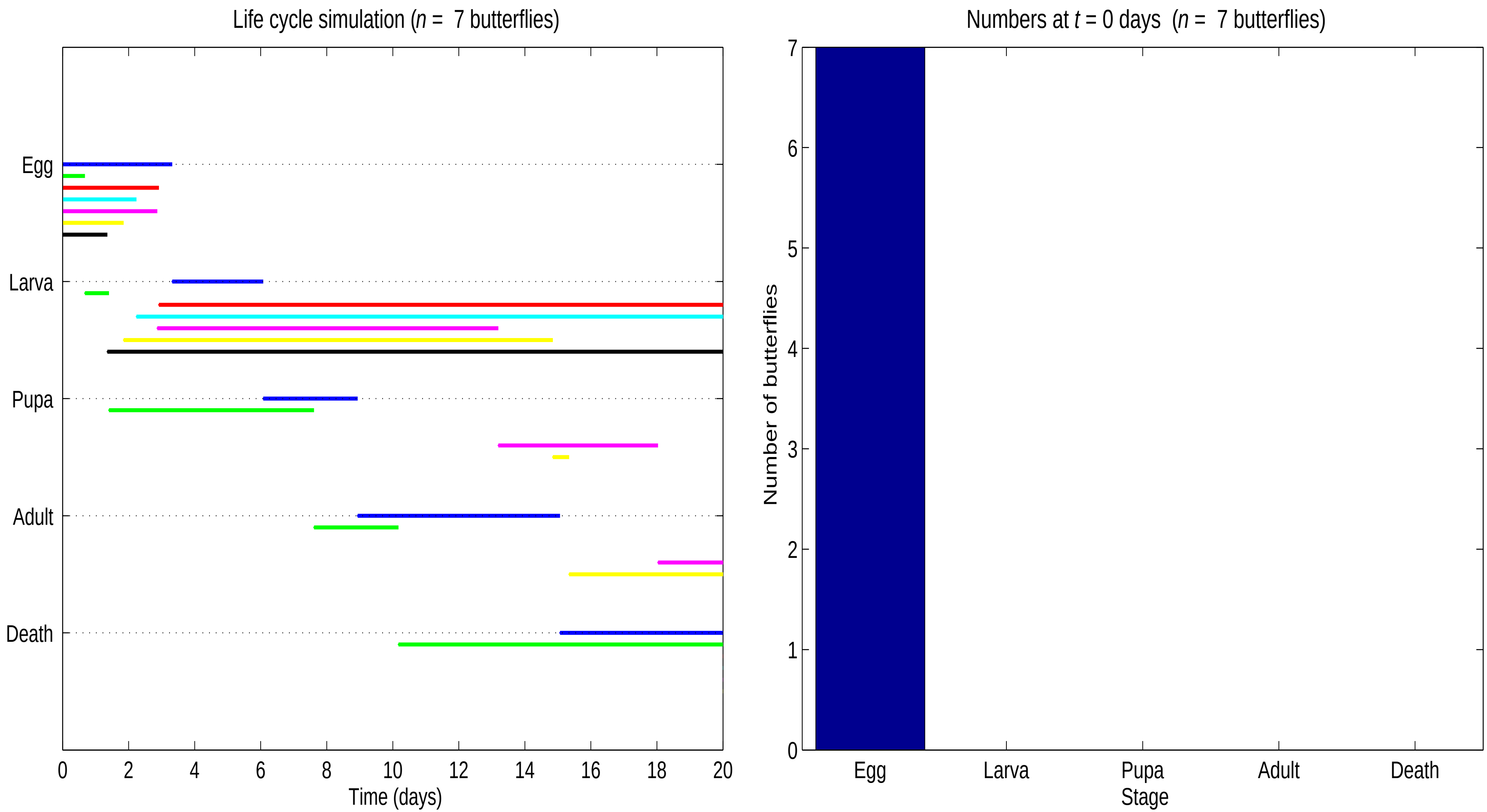
$$q(\boldsymbol{n}, \boldsymbol{n} + \boldsymbol{e}_j - \boldsymbol{e}_i) = n_i q_{ij},$$

for all states $j \neq i$ in $S$, where $\boldsymbol{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ is the unit vector with a $1$ as its $j$-th entry (this transition corresponds to a single individual moving from state $i$ to state $j$).

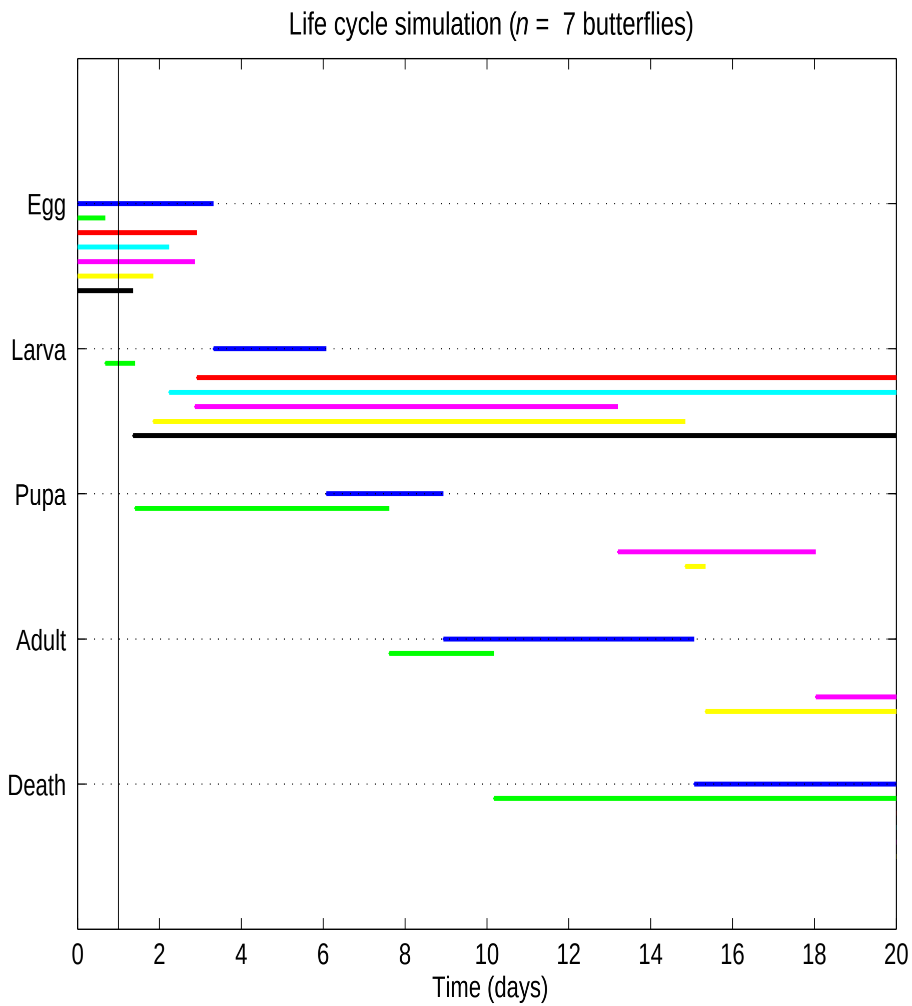
# Ensemble of organisms

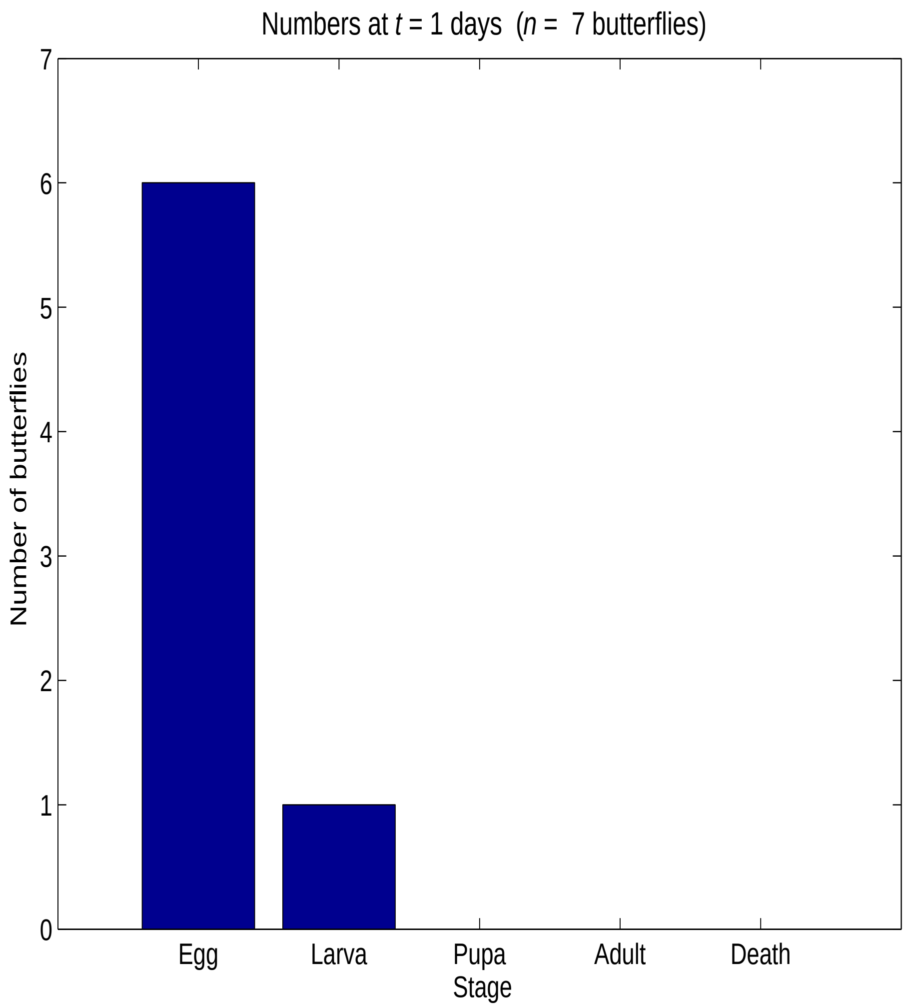


# Ensemble state description

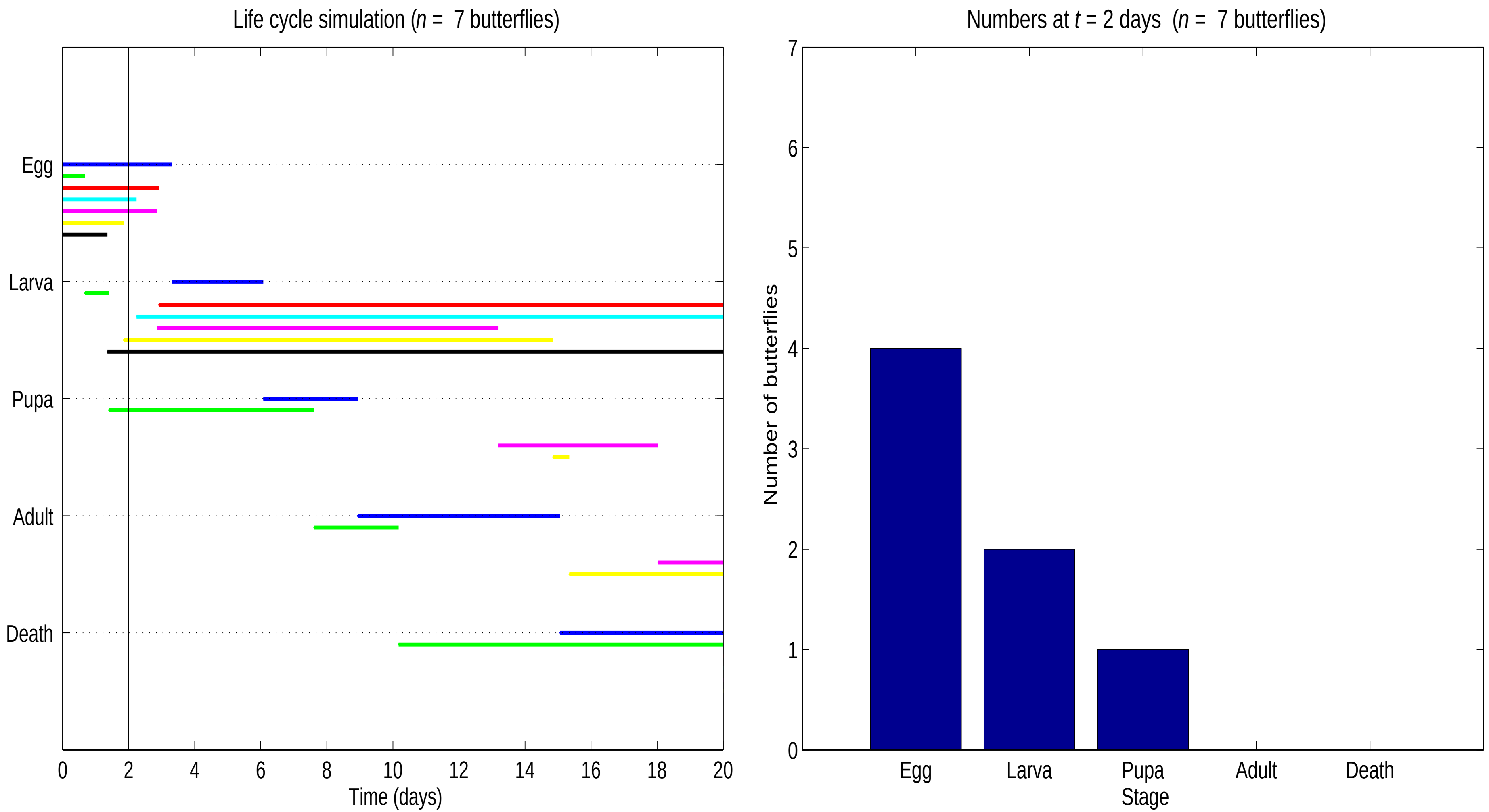


# Ensemble state description

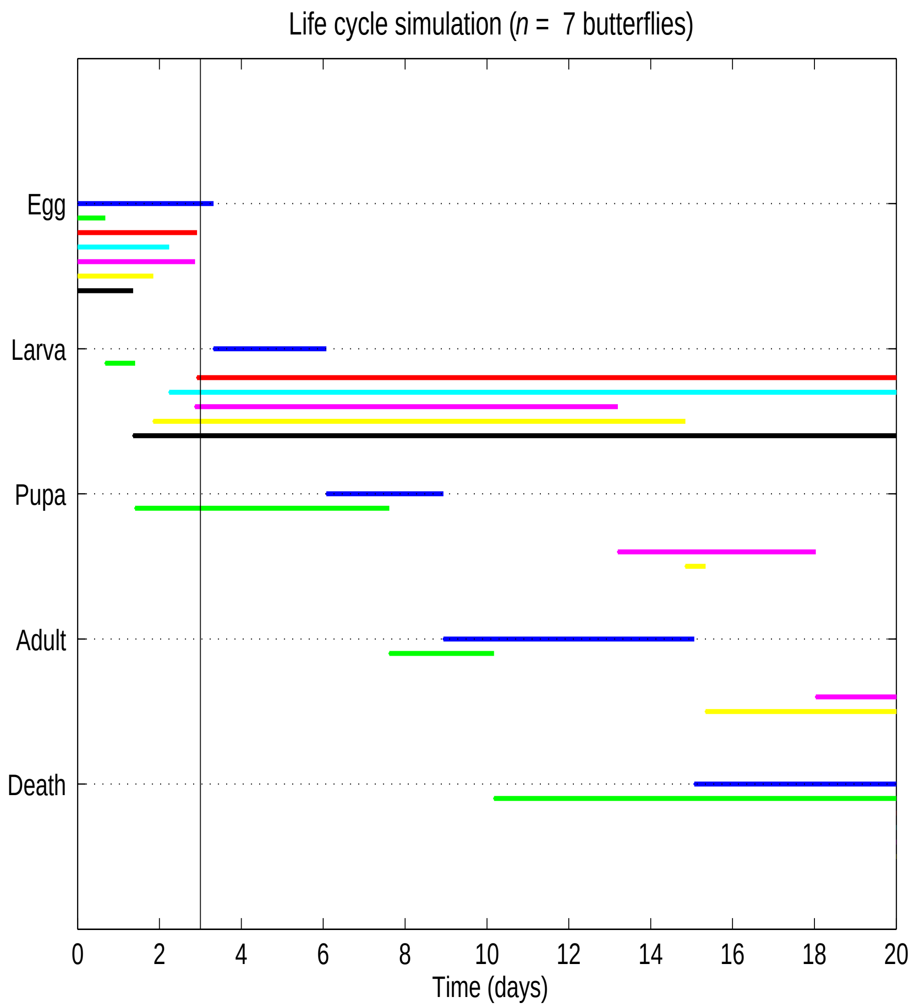


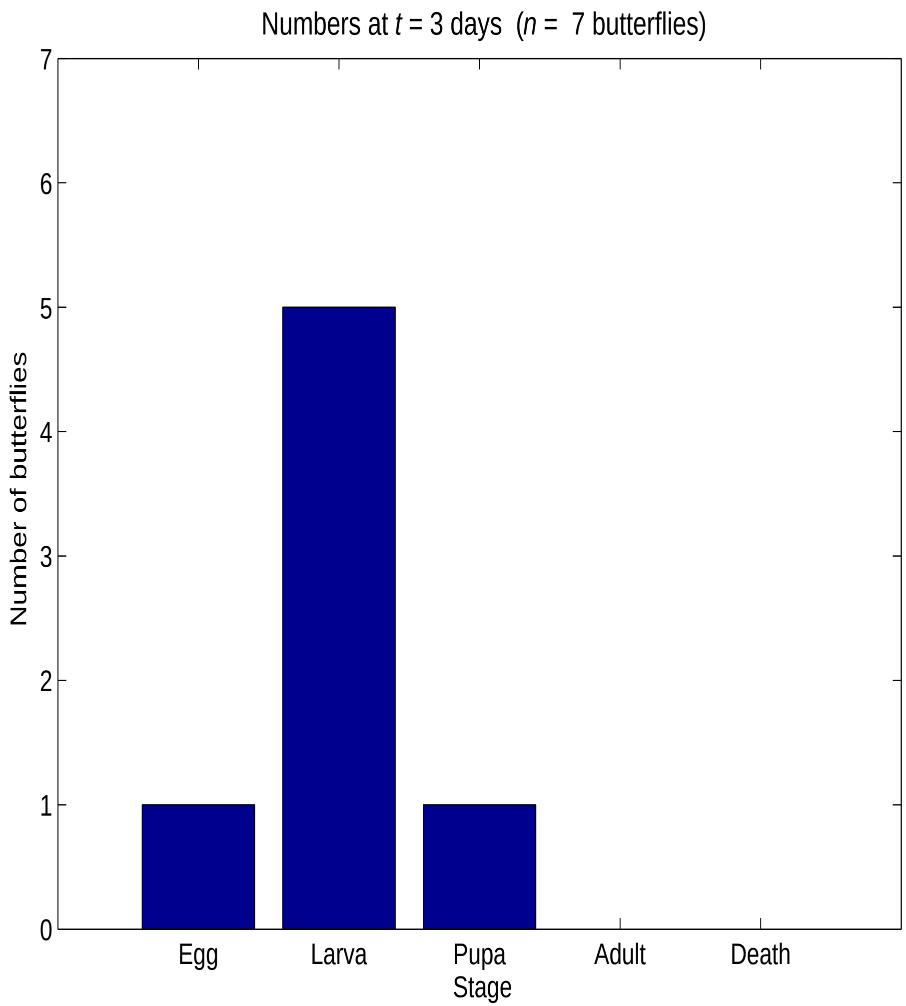



# Ensemble state description

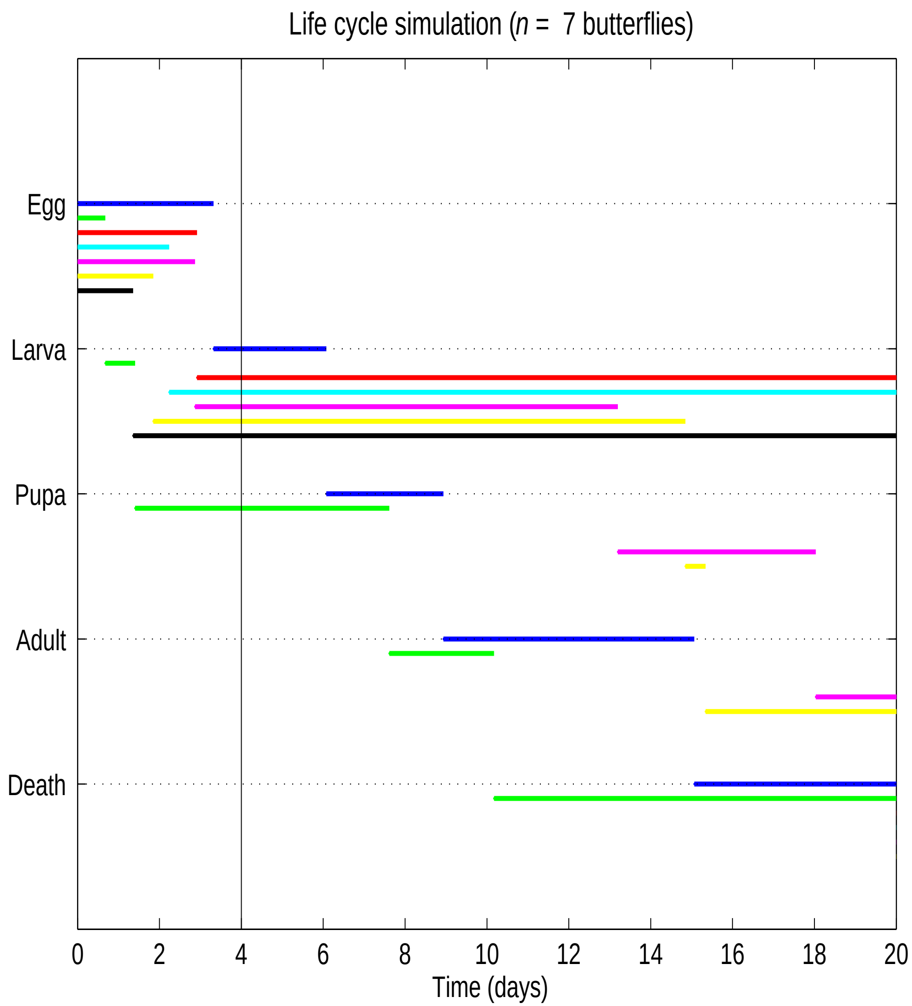


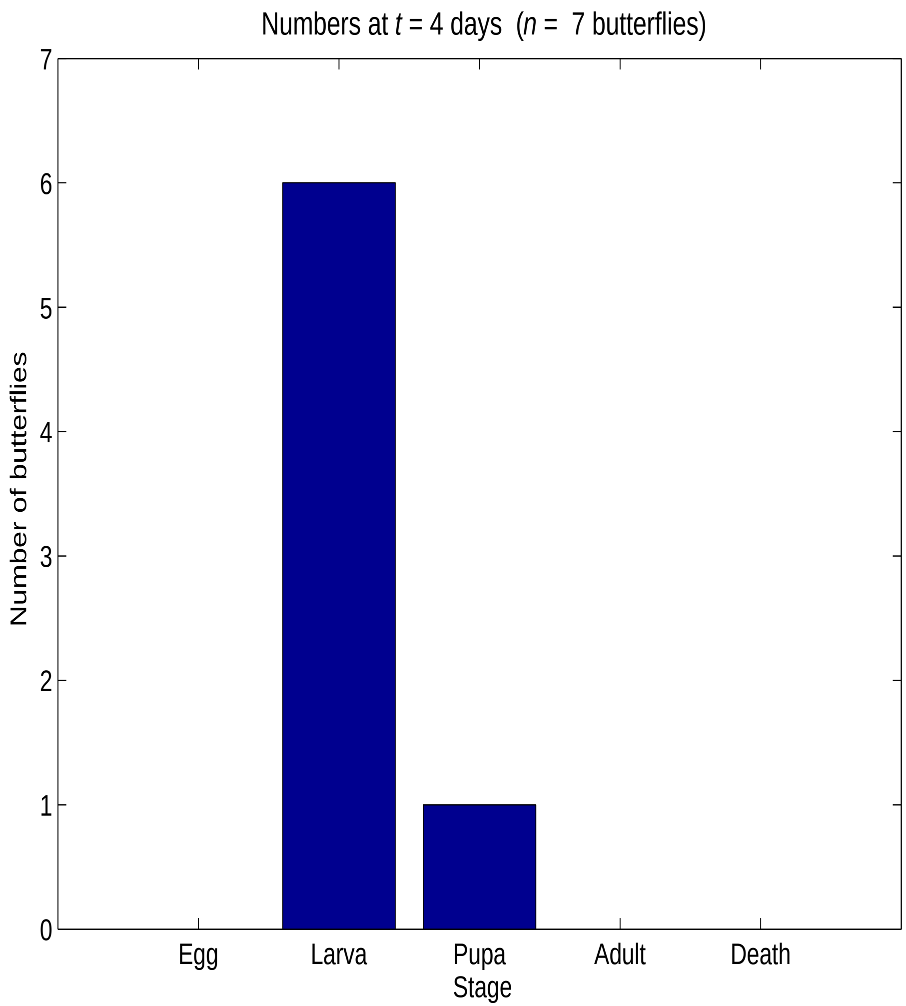
# Ensemble state description



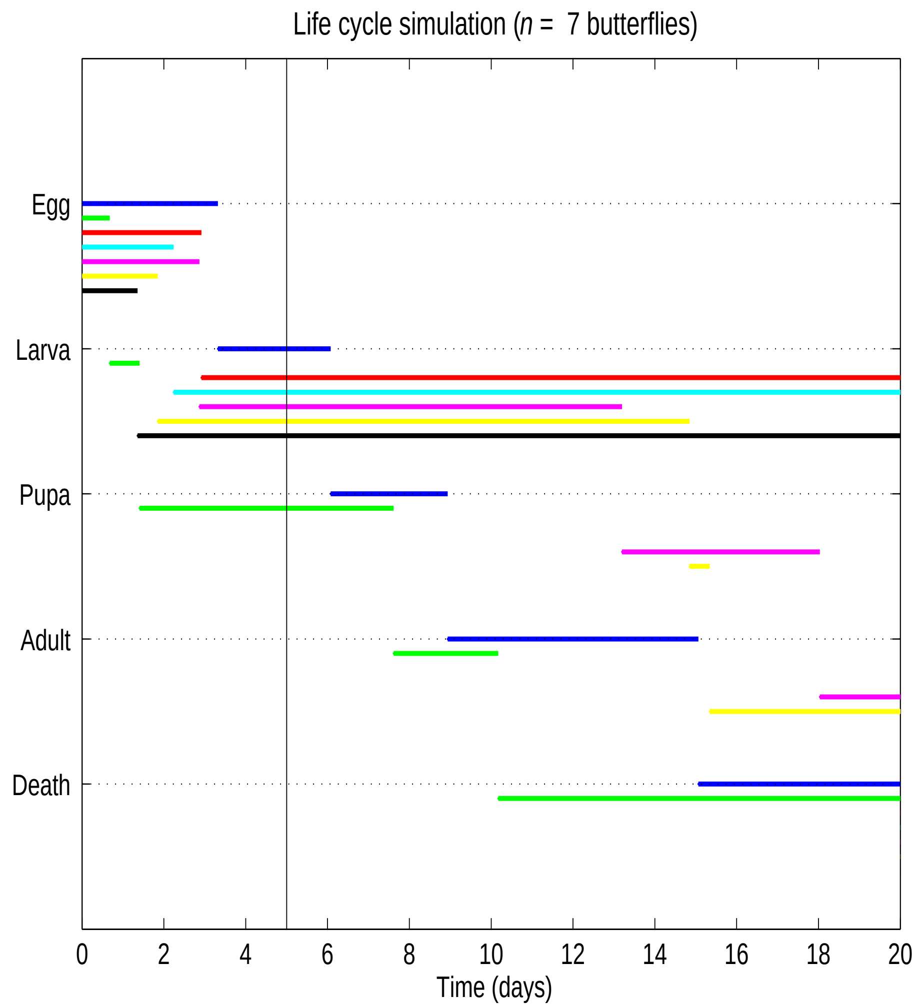



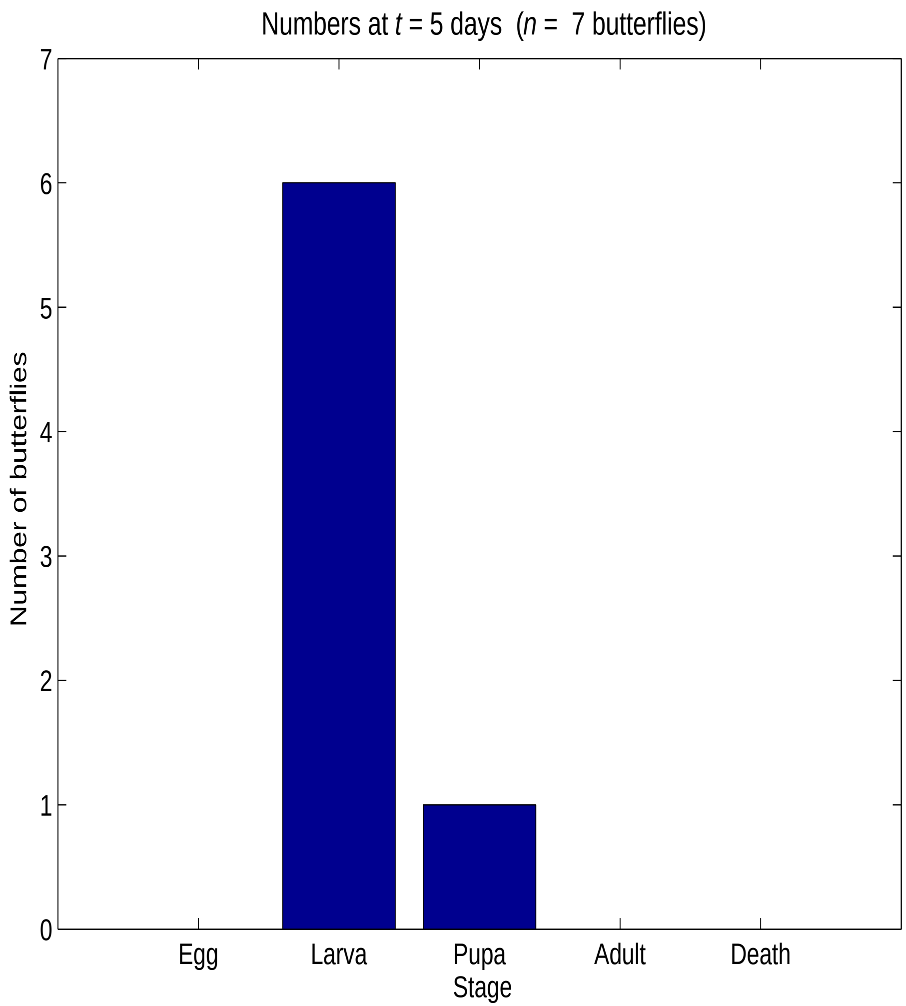
# Ensemble state description



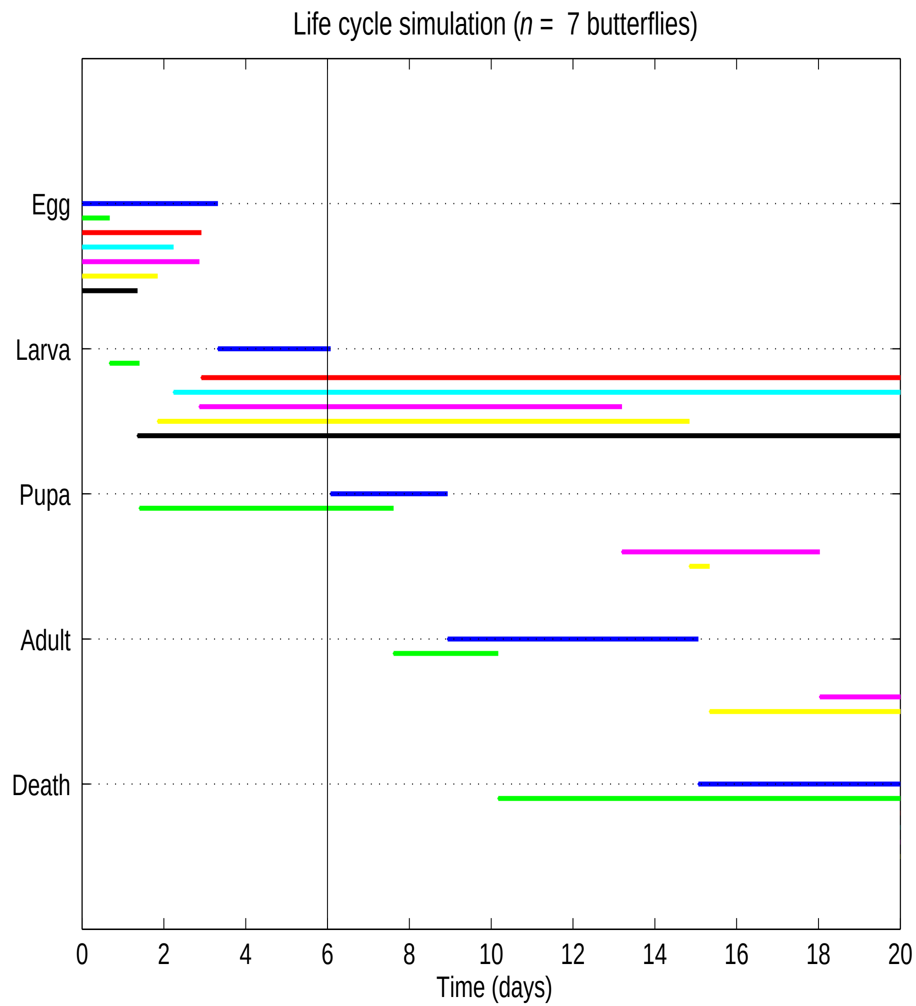



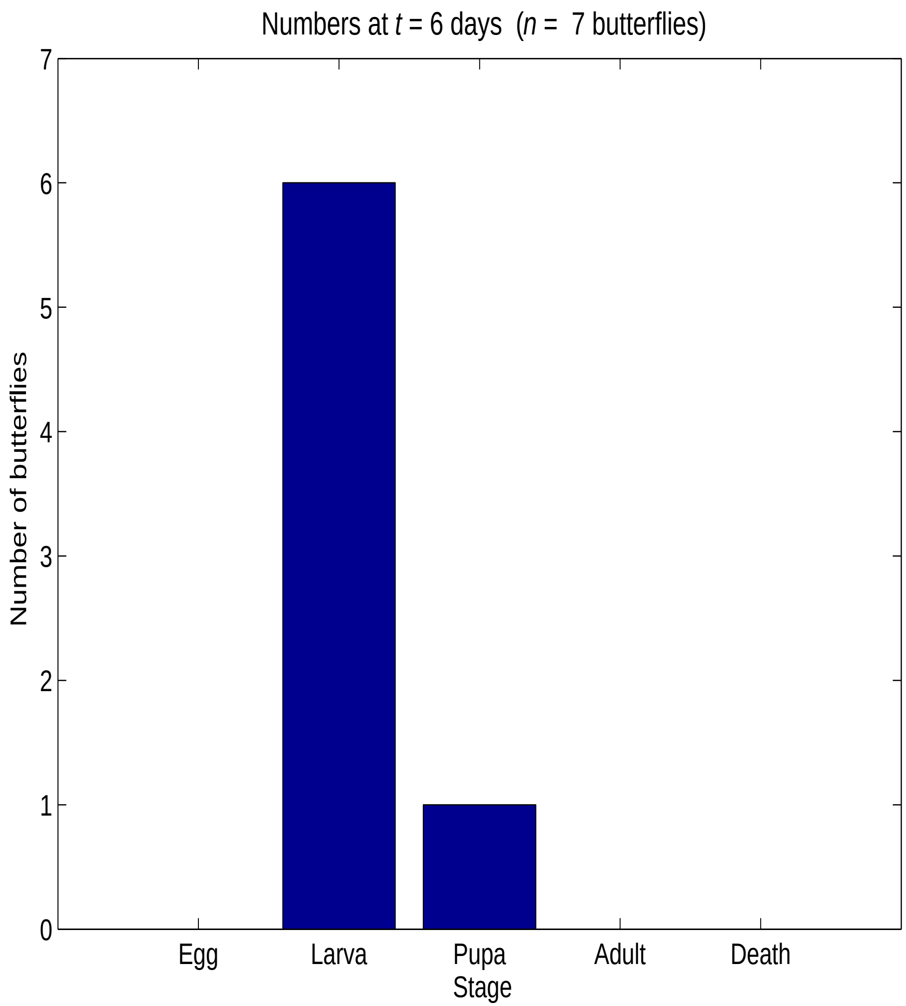
# Ensemble state description



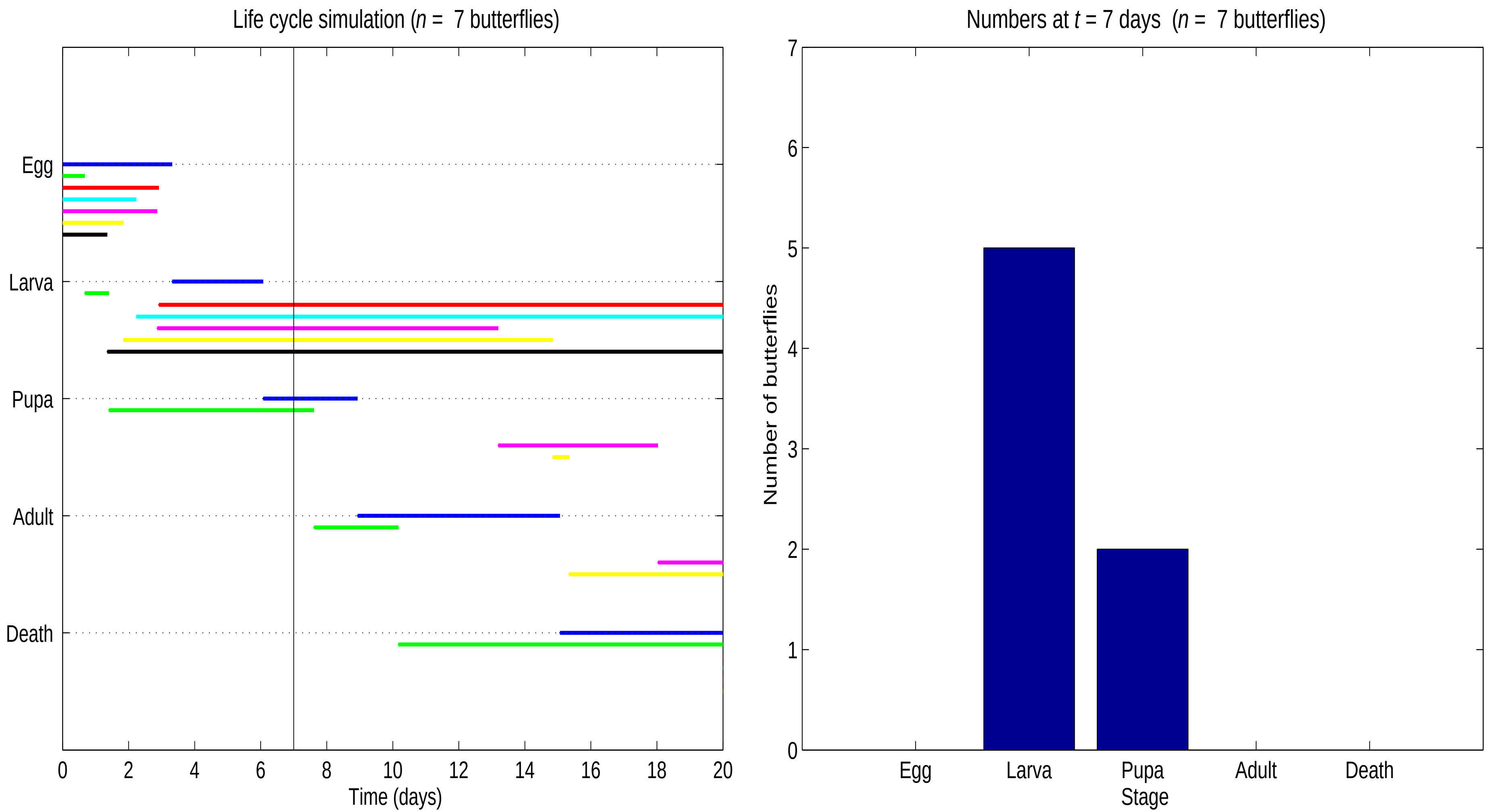



# Ensemble state description

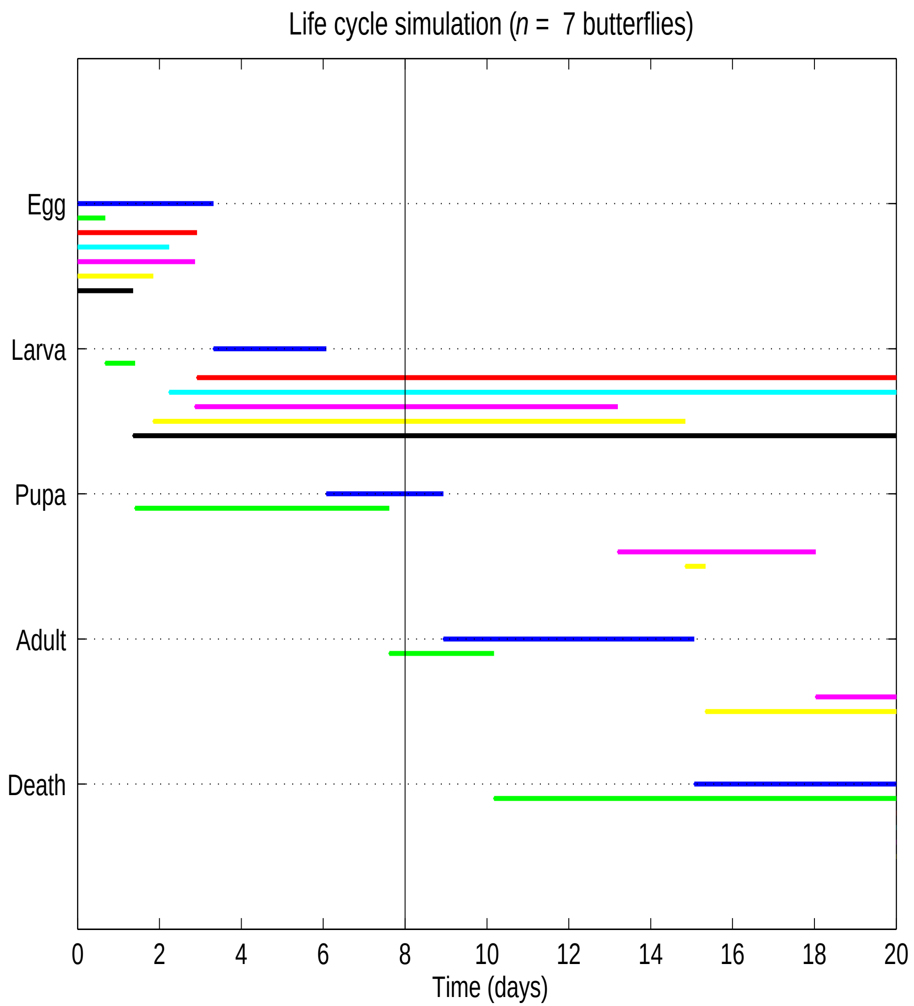


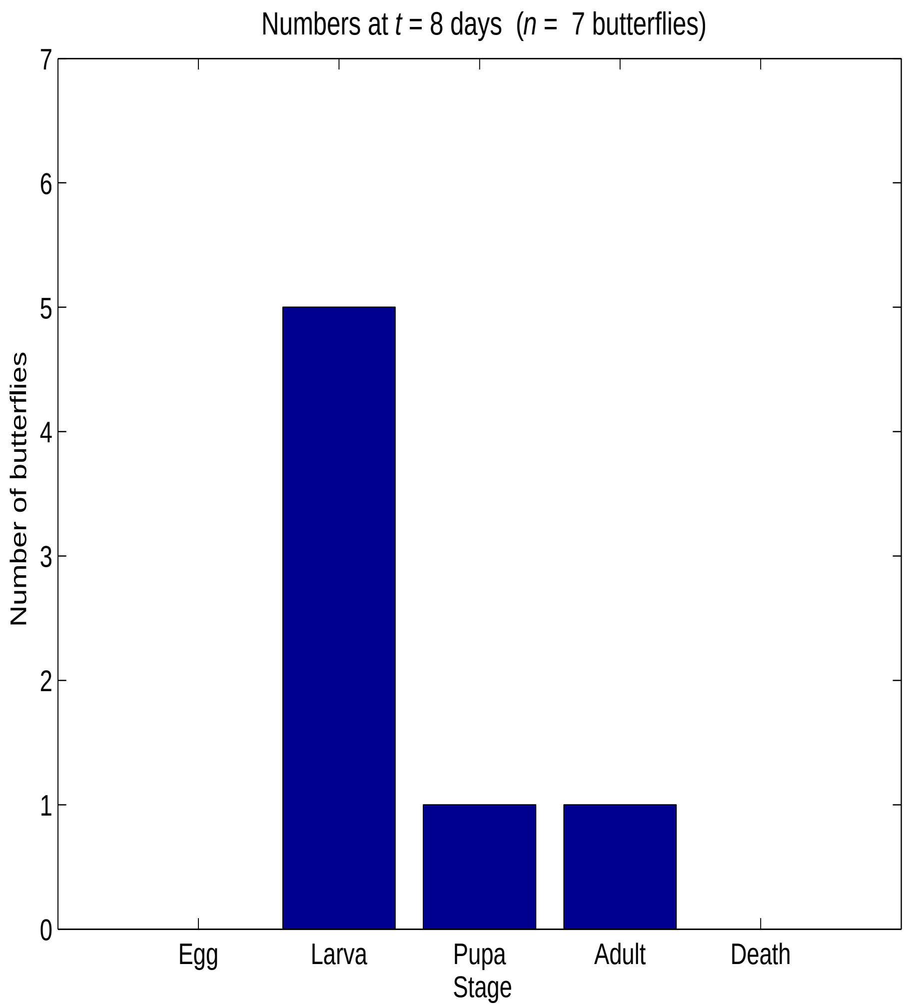



# Ensemble state description

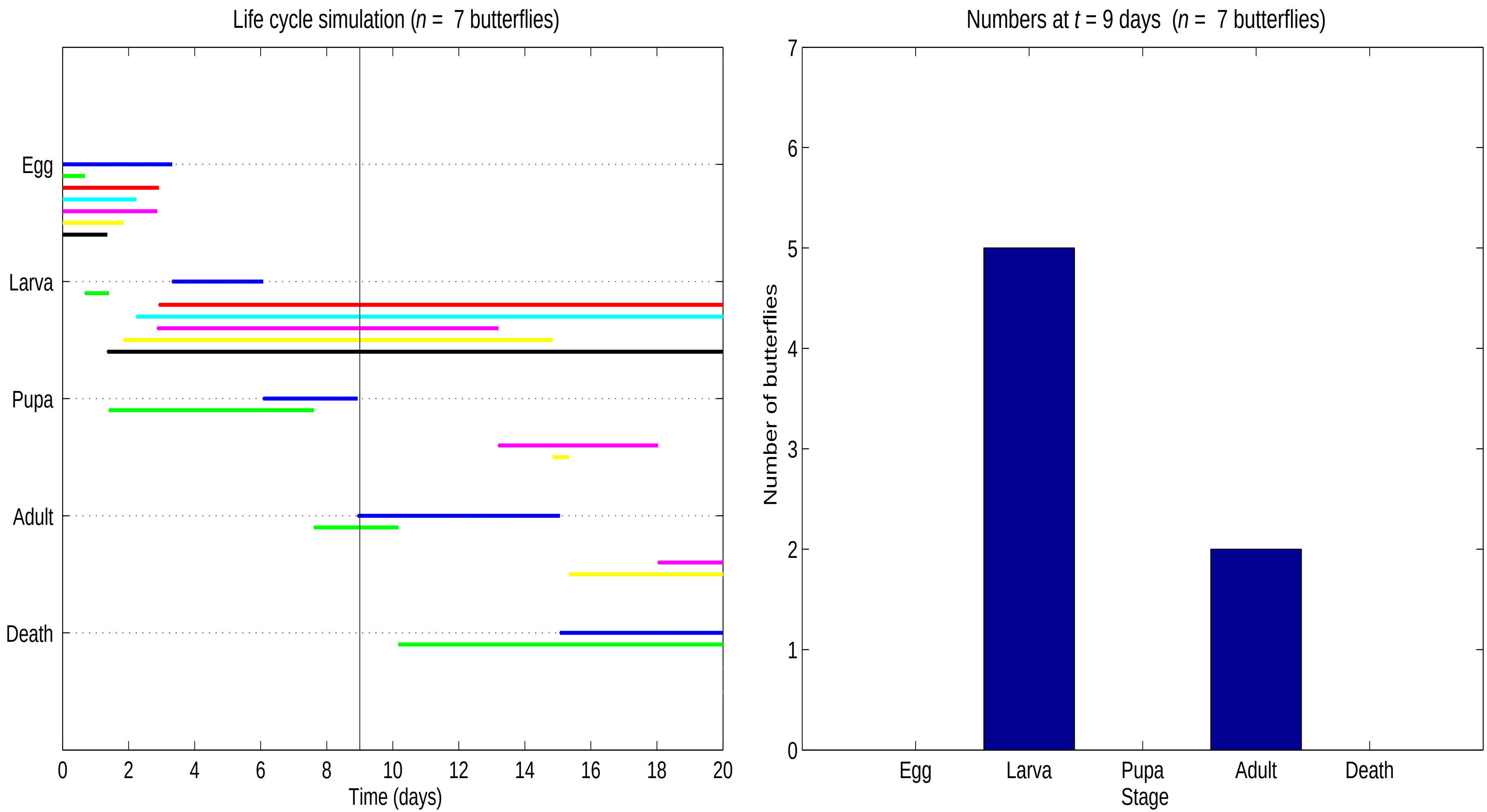


# Ensemble state description

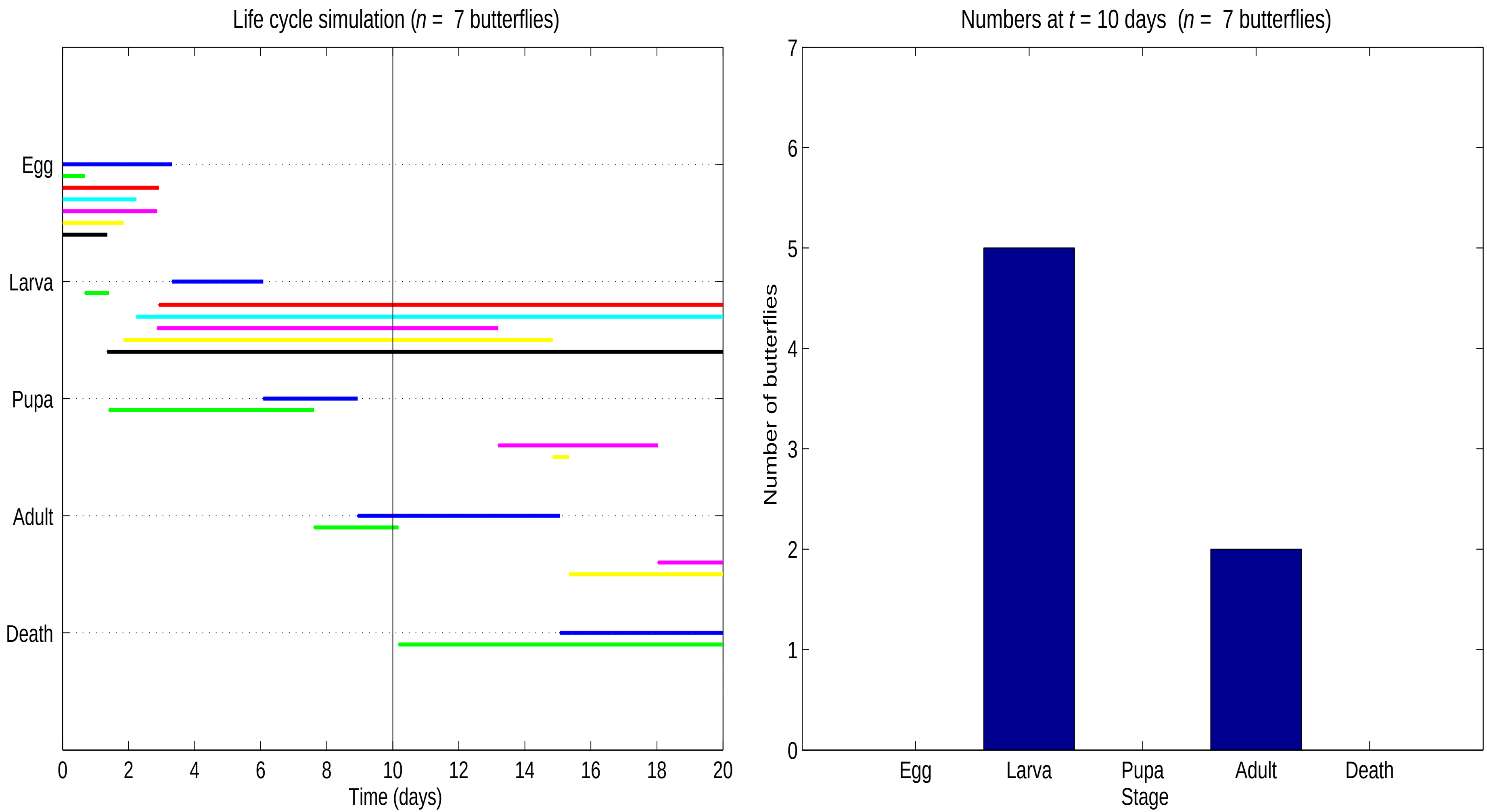


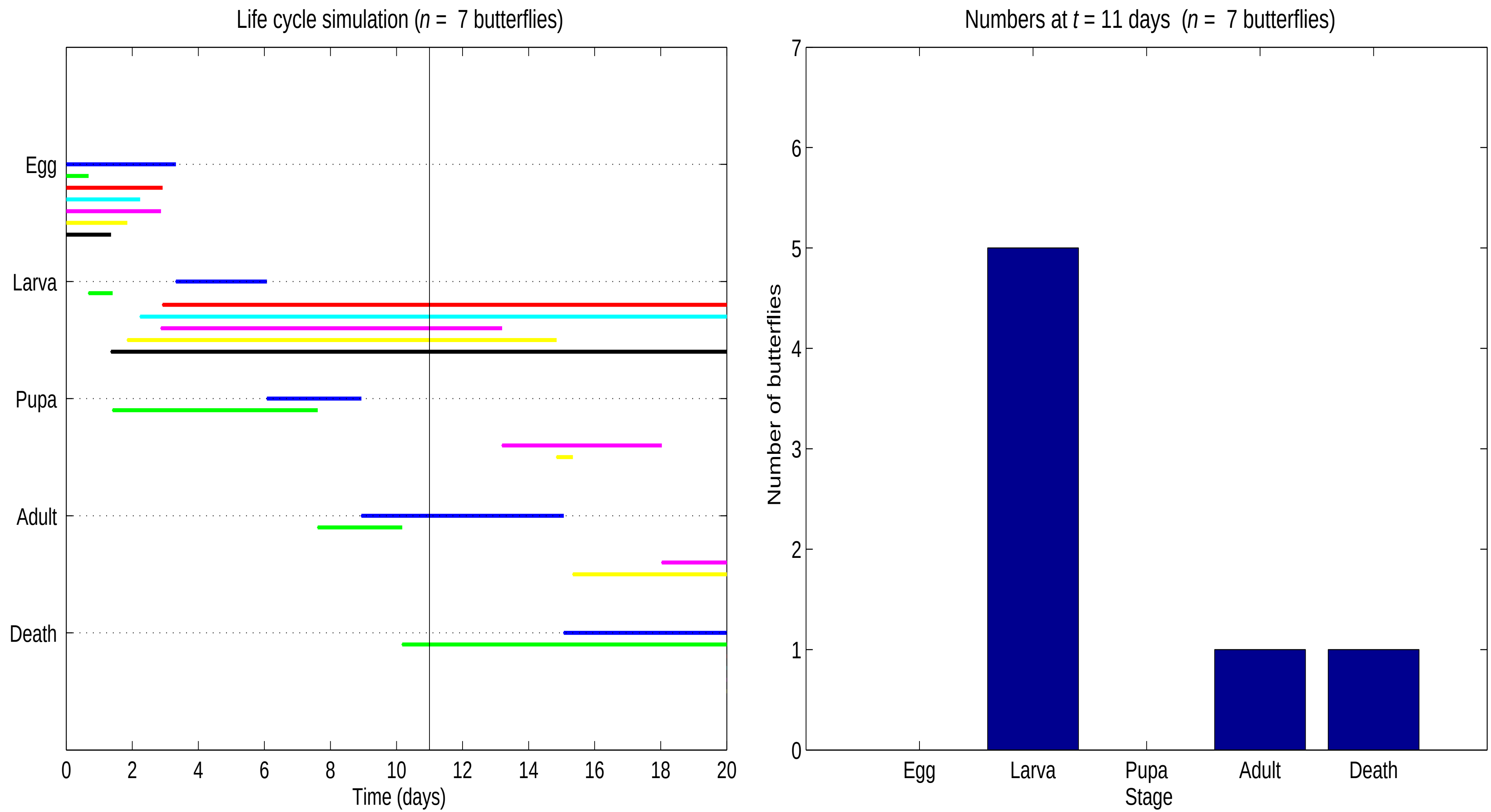
# Ensemble state description



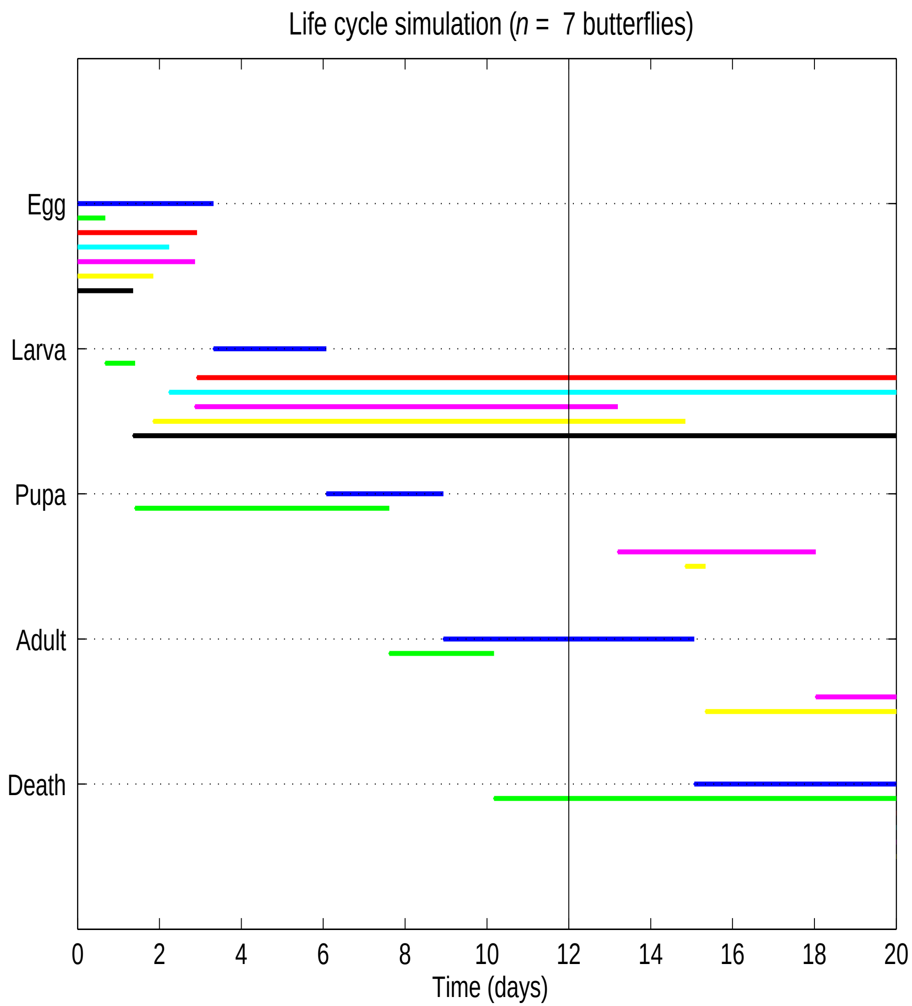
# Ensemble state description

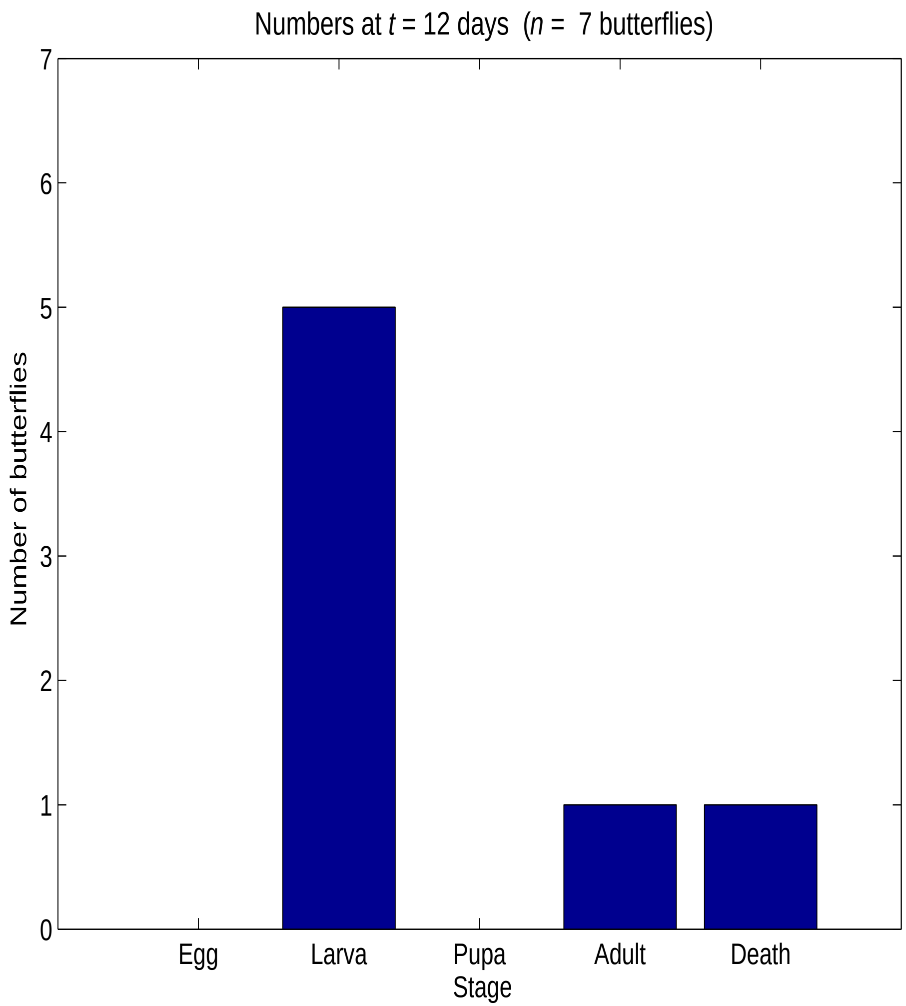


# Ensemble state description

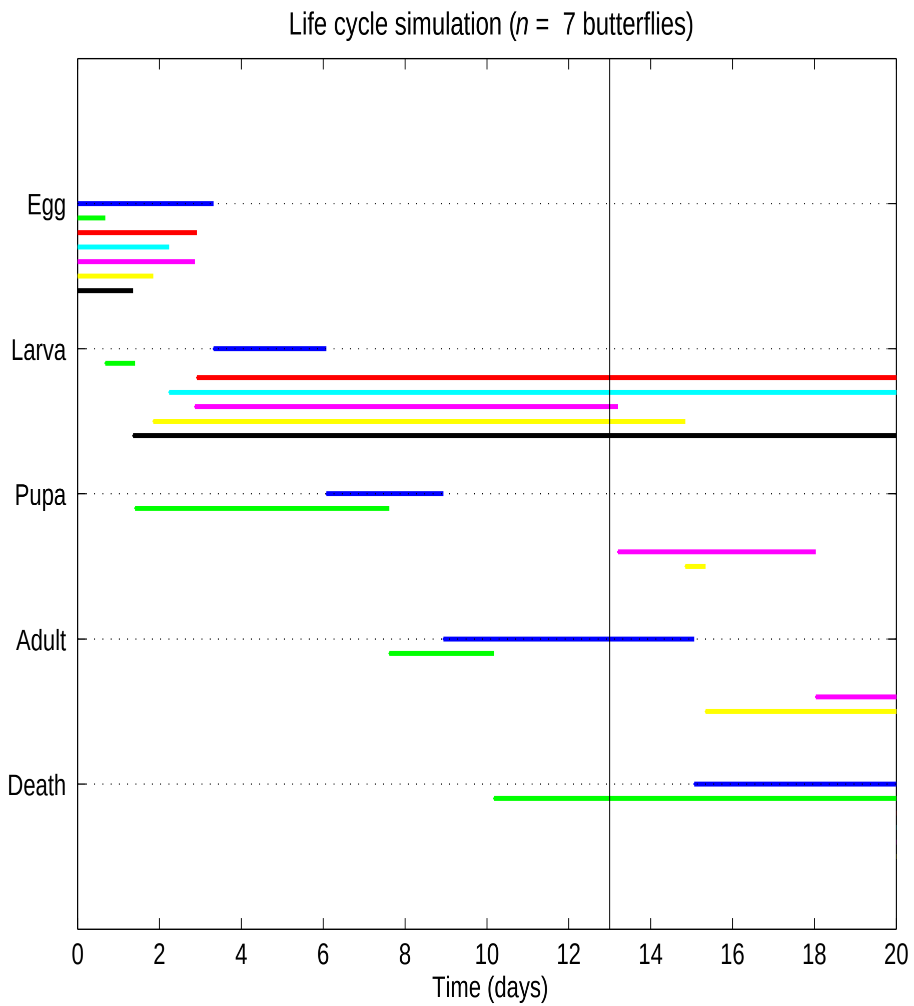


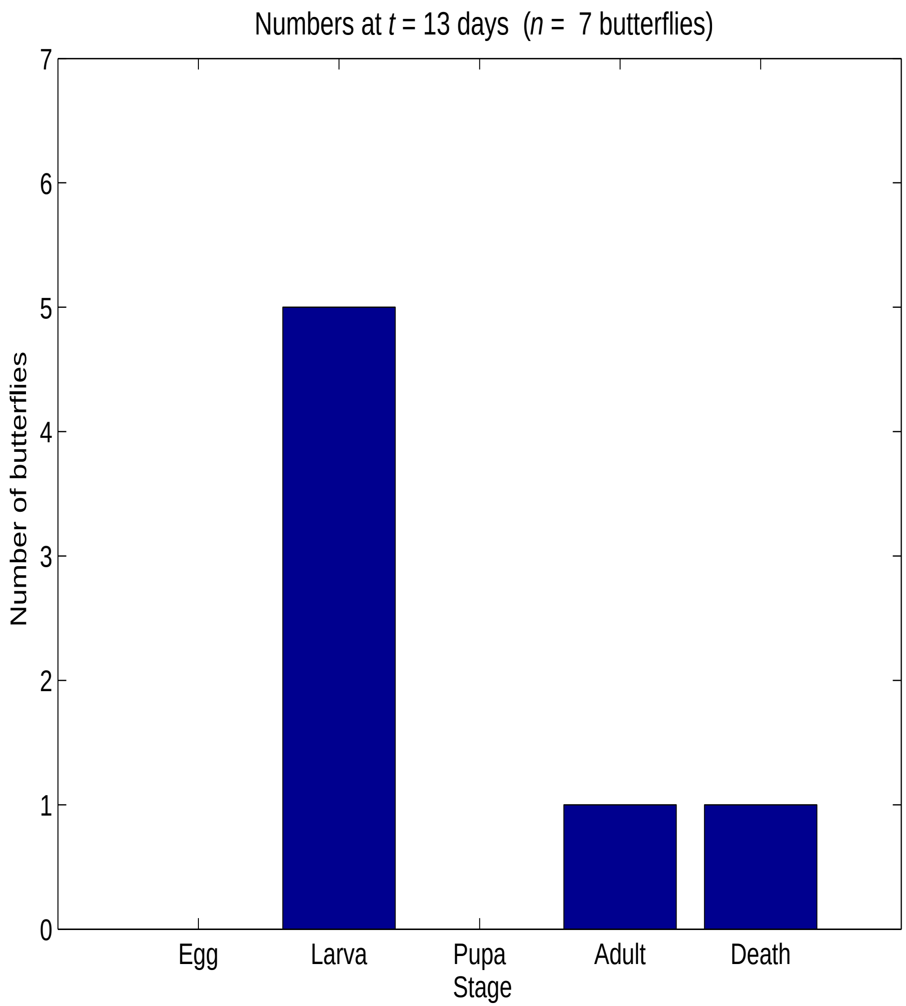
# Ensemble state description



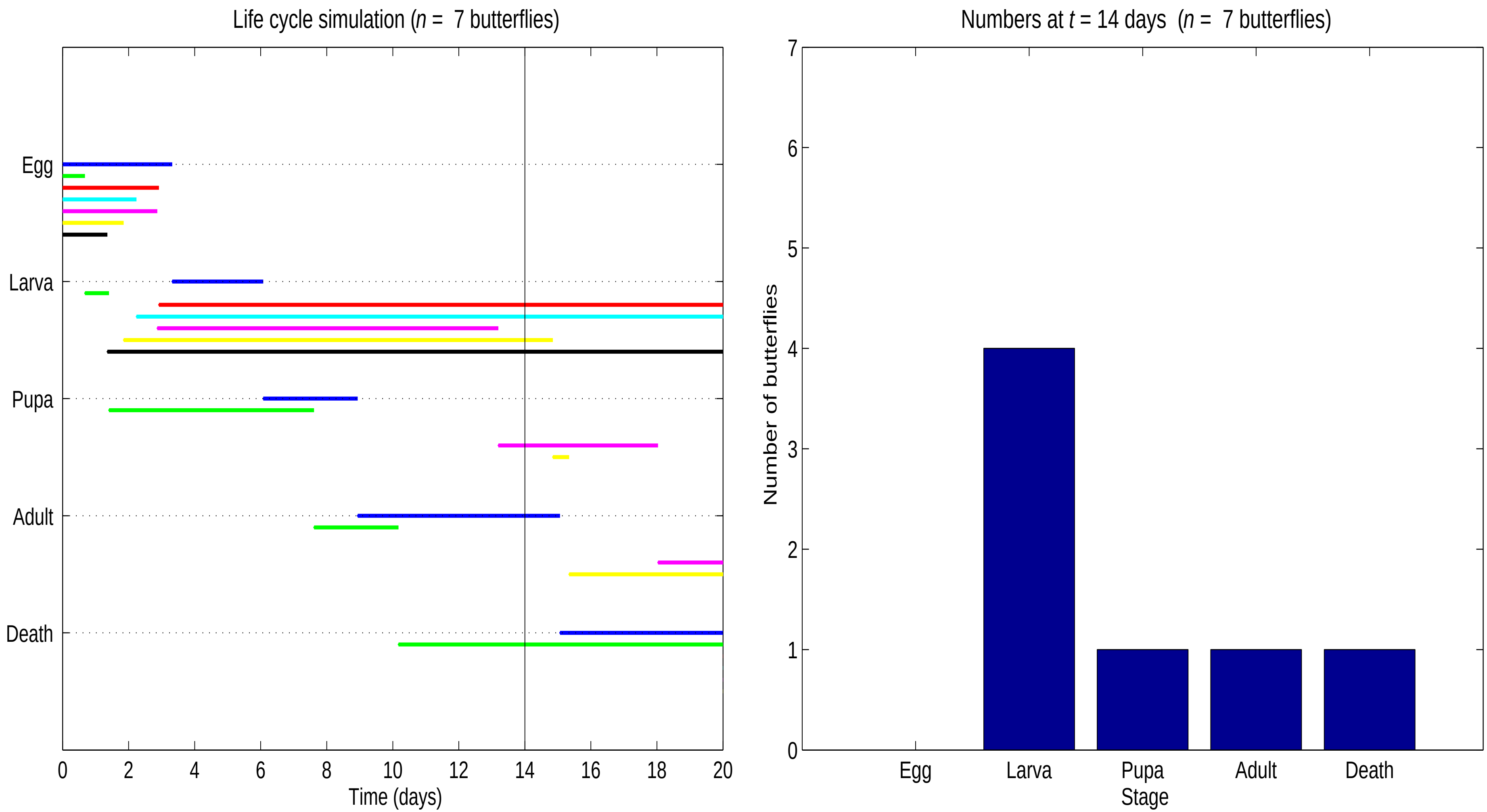



# Ensemble state description

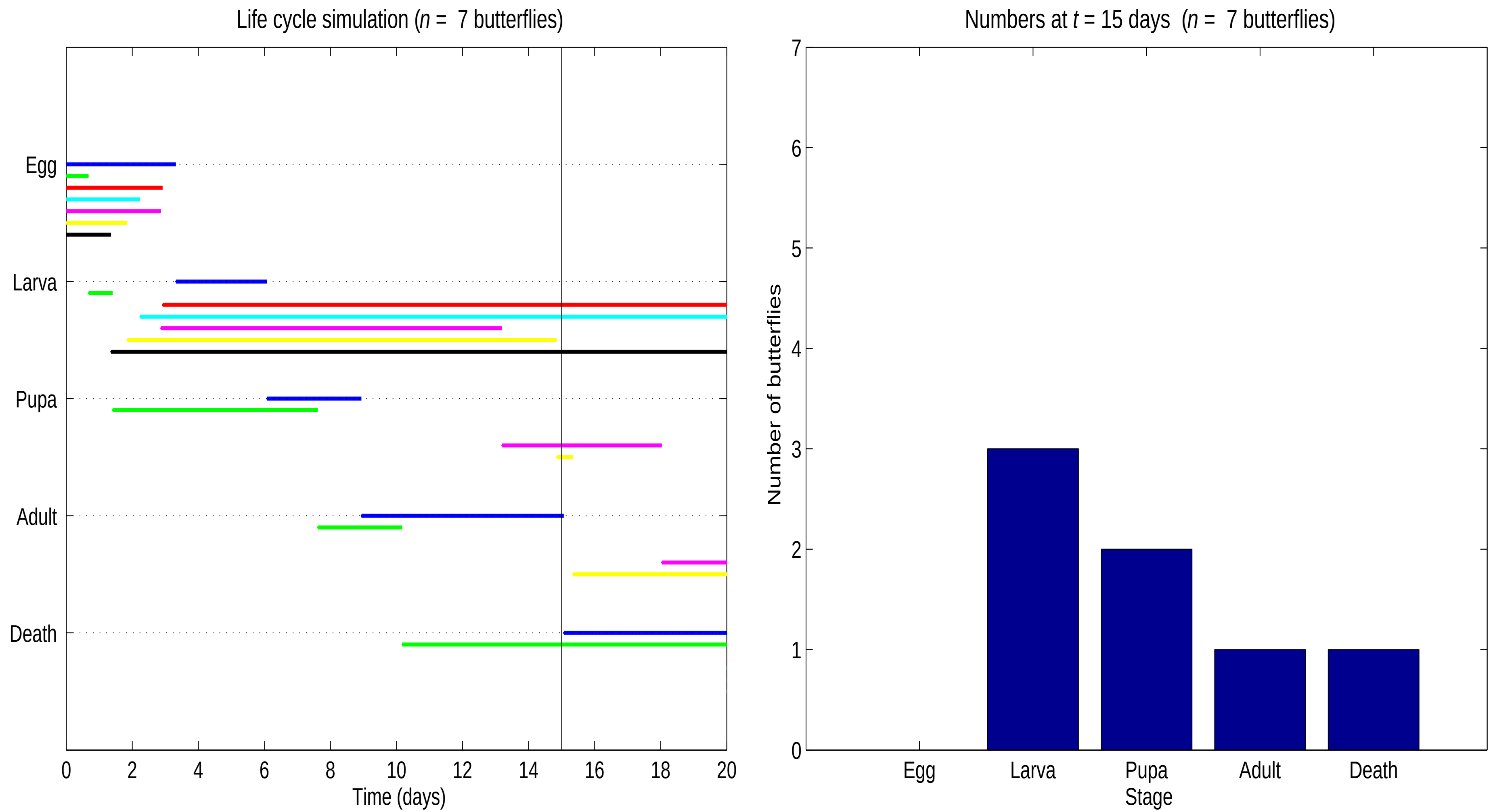


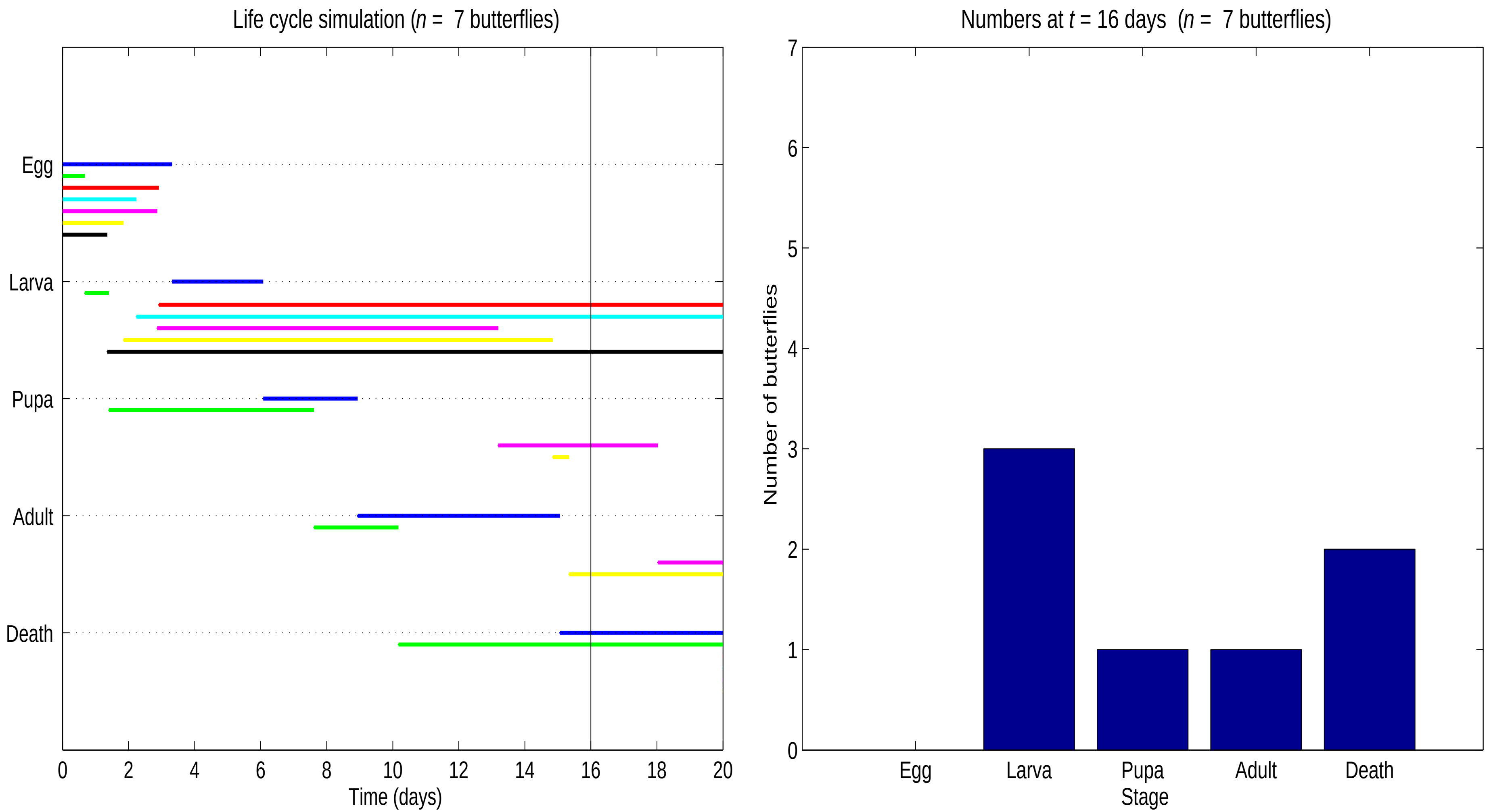
# Ensemble state description



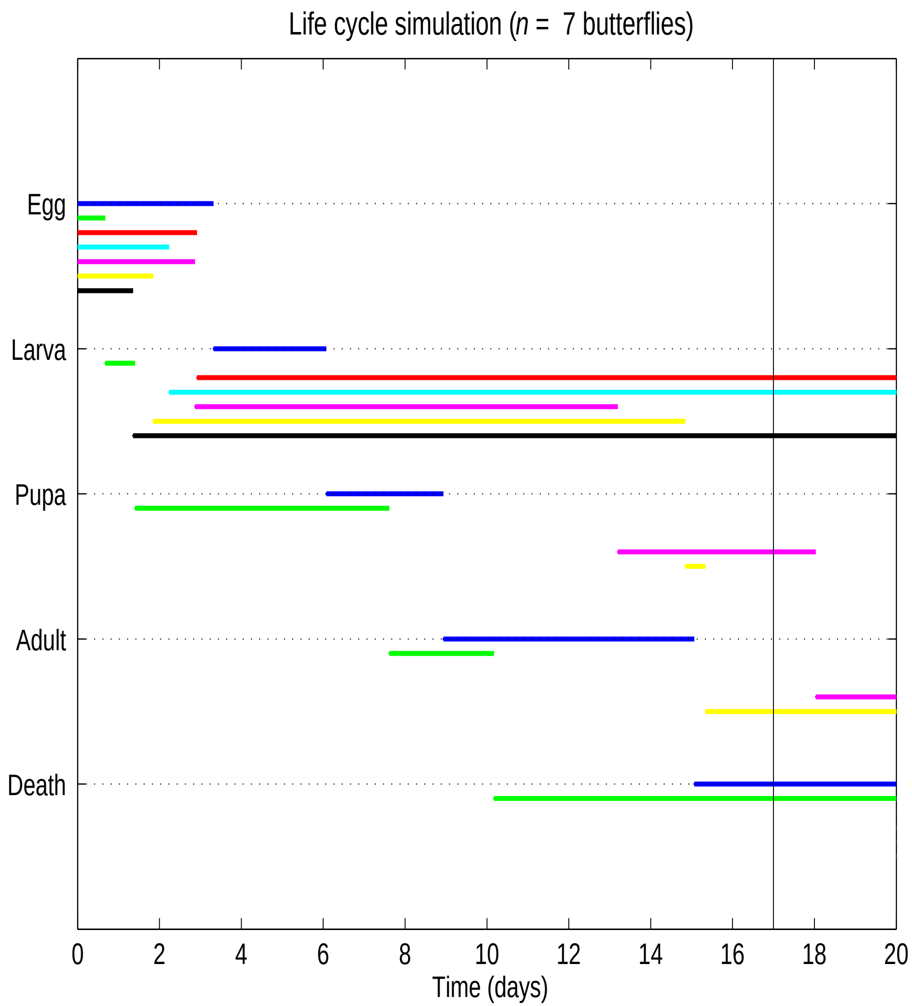
# Ensemble state description

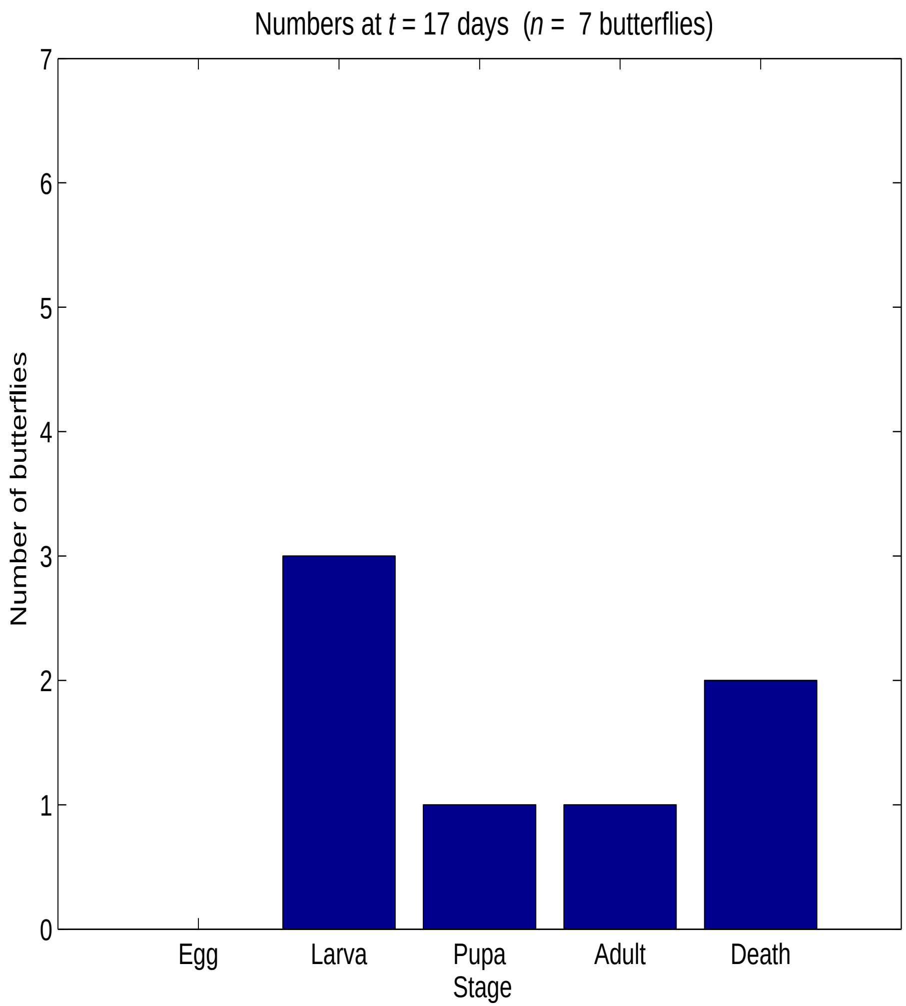


# Ensemble state description

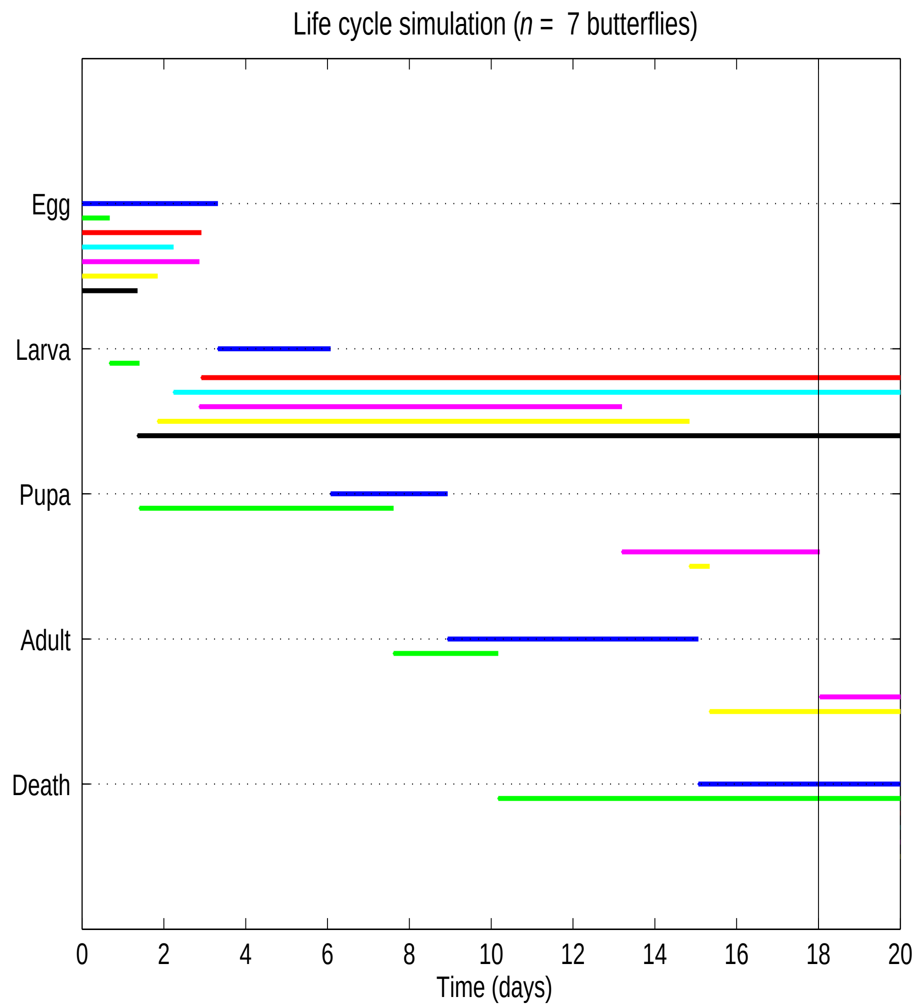


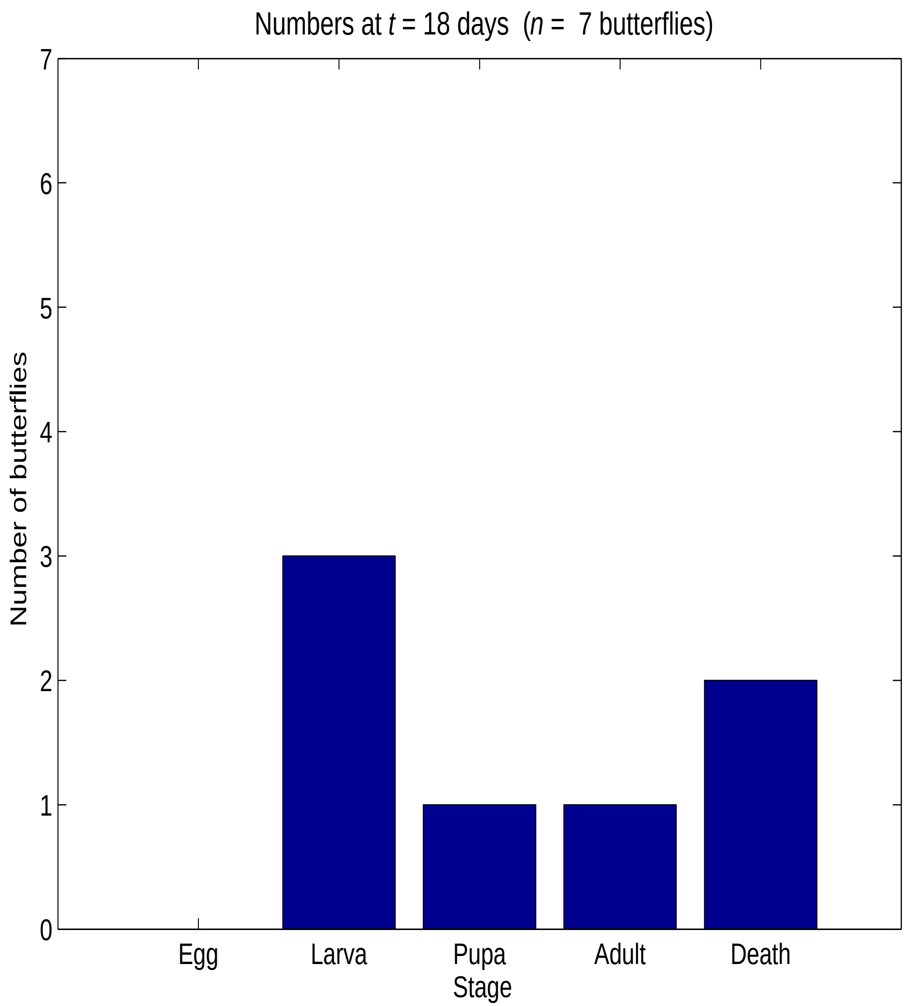
# Ensemble state description



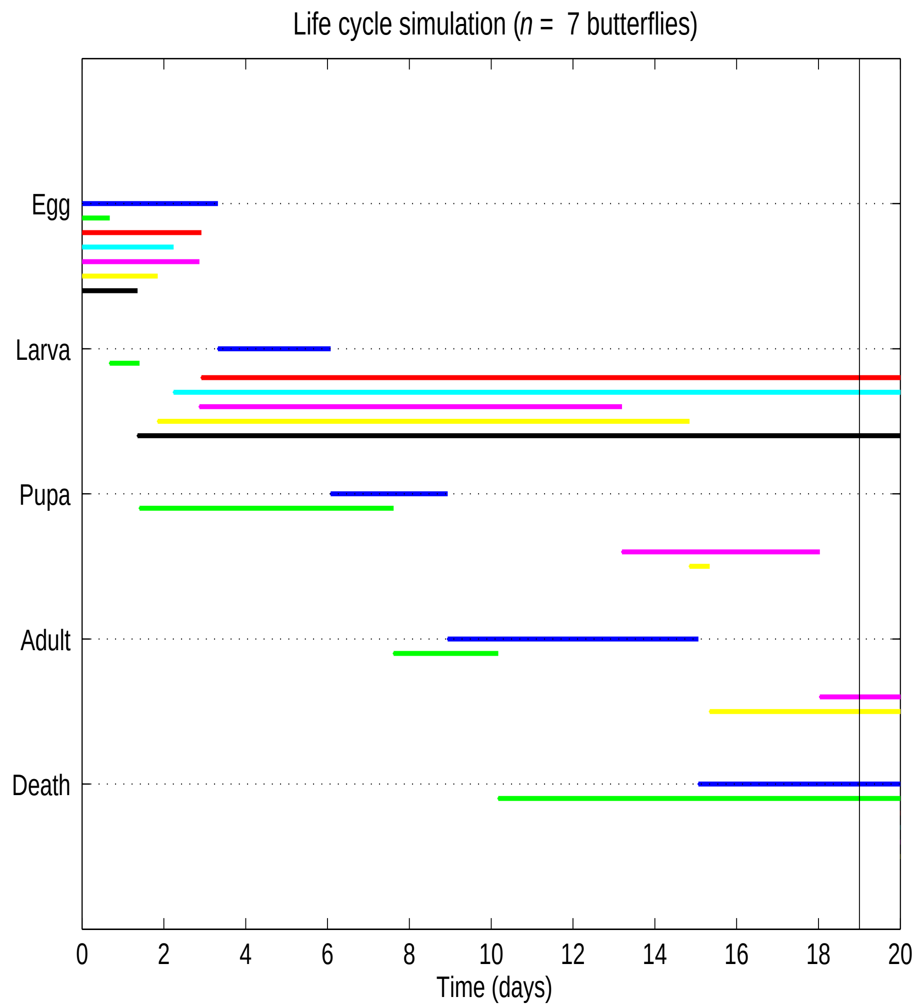



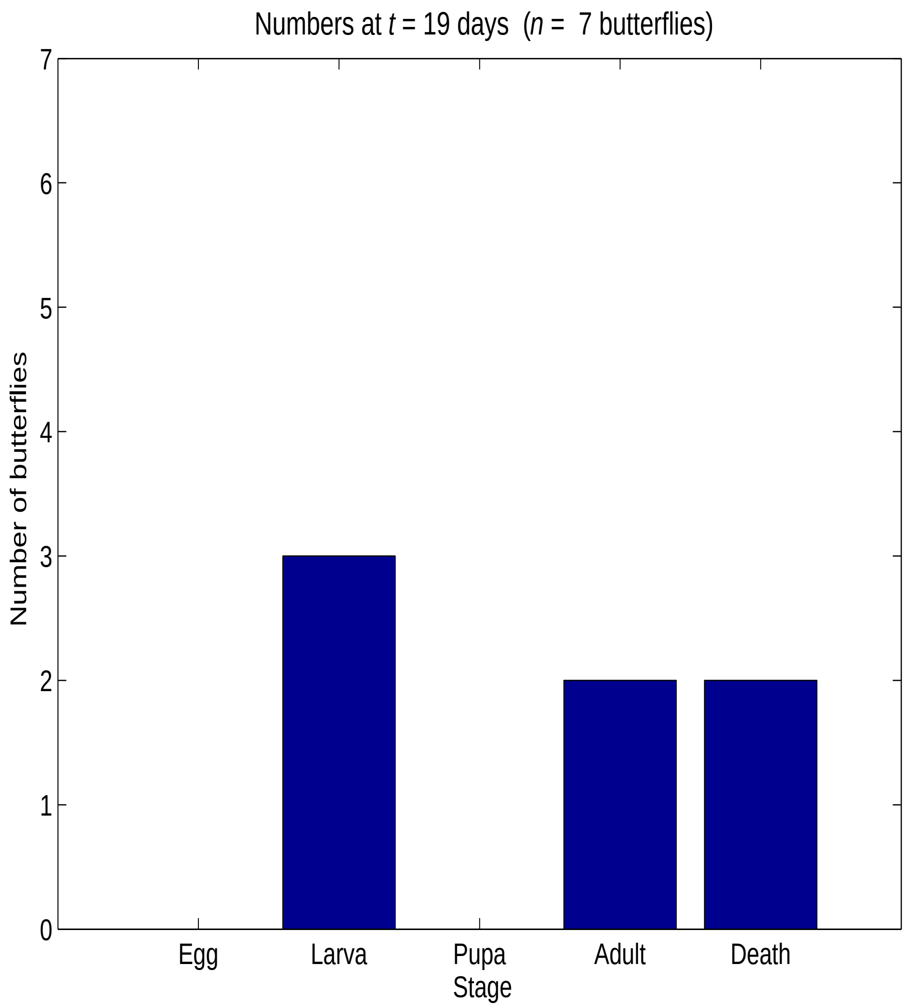
# Ensemble state description



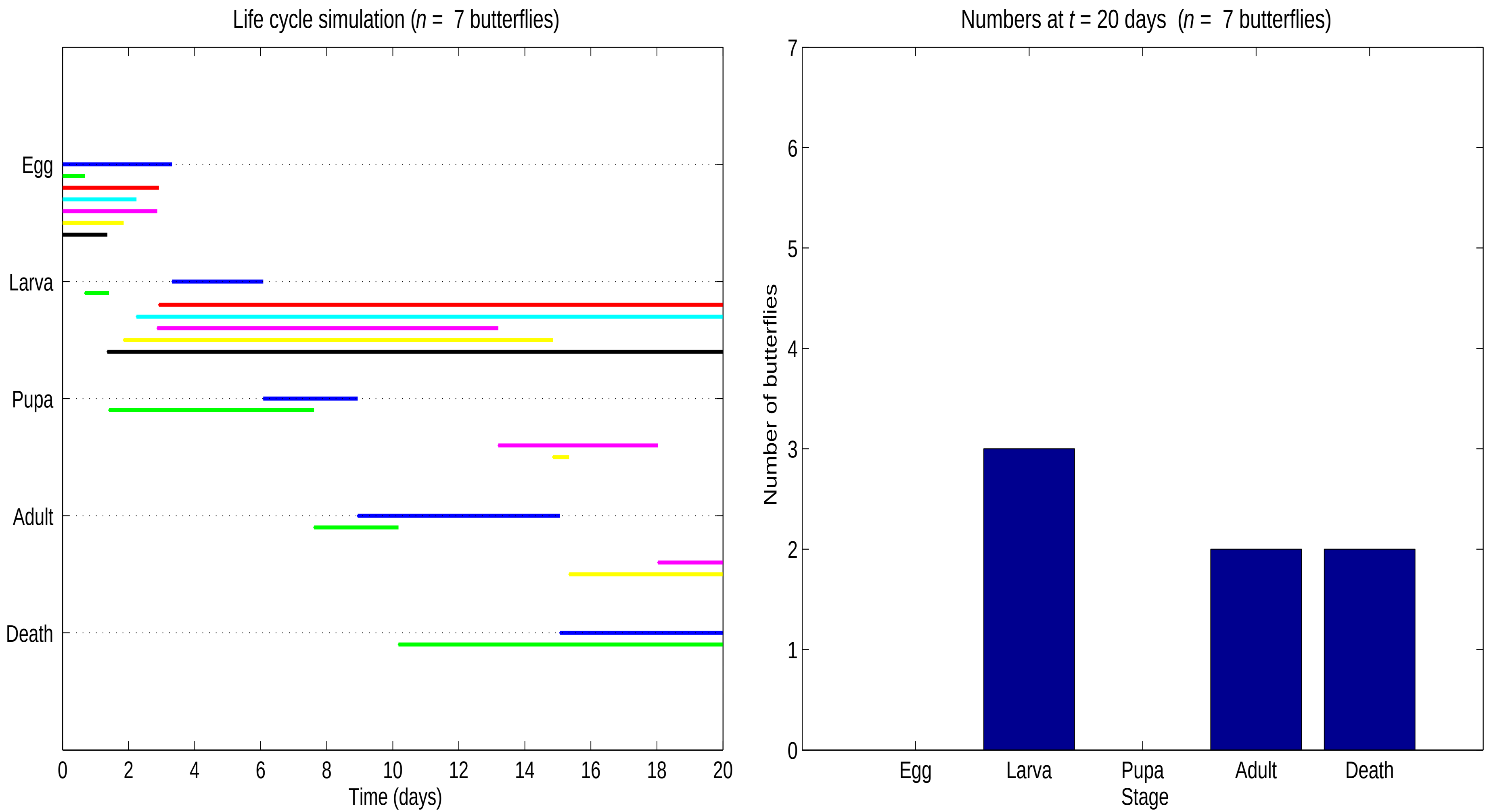



# Ensemble state description

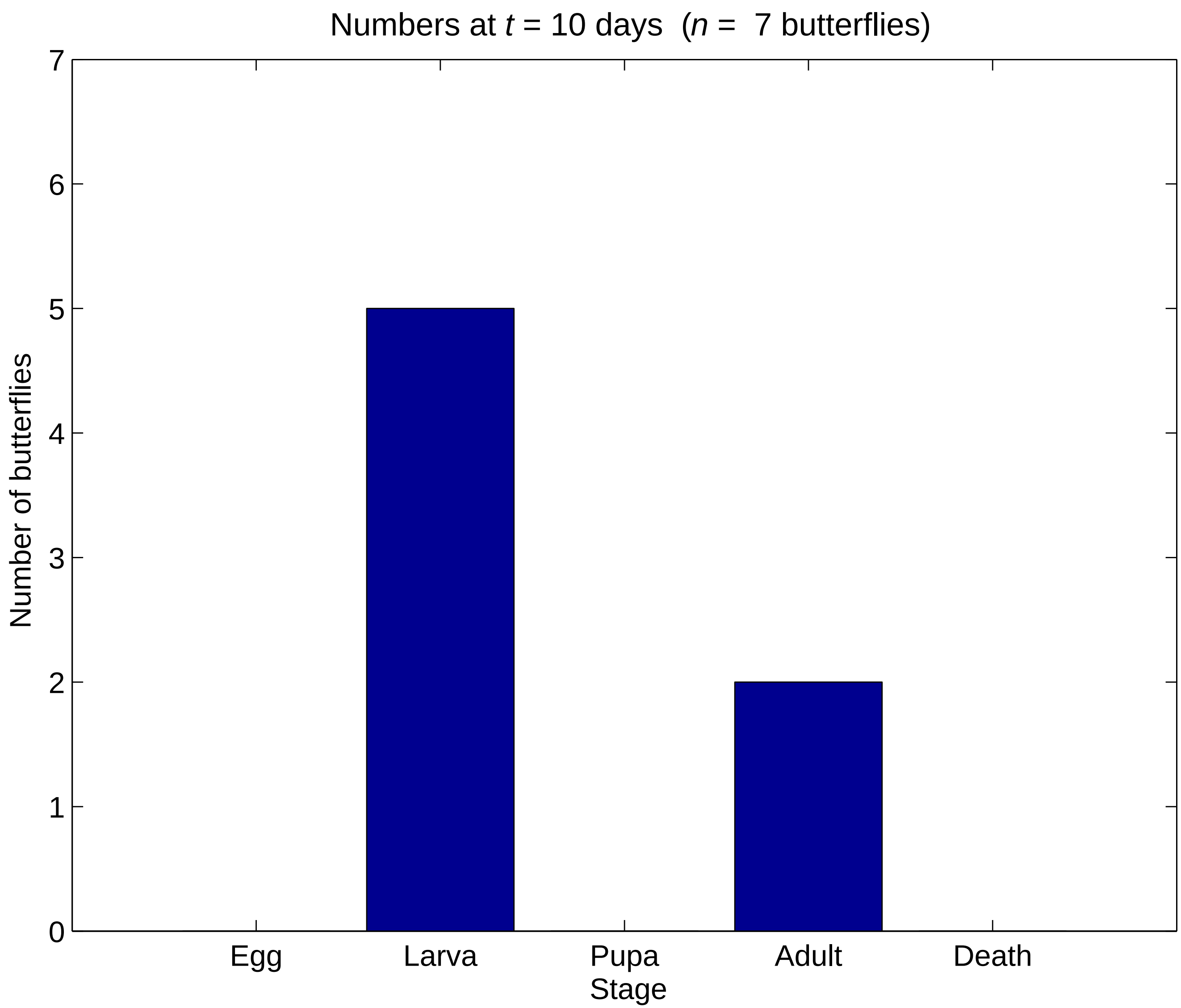


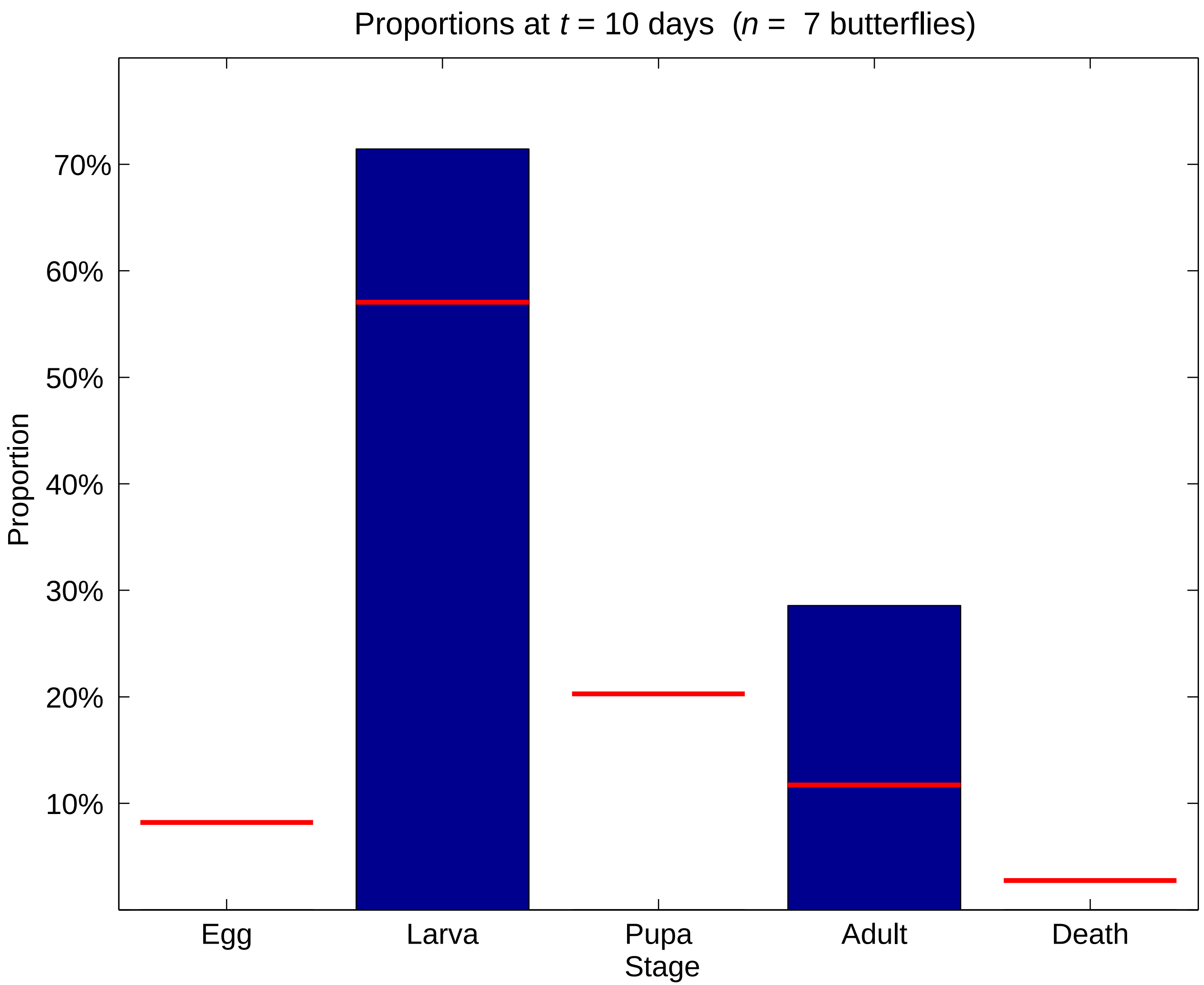
# Ensemble state description



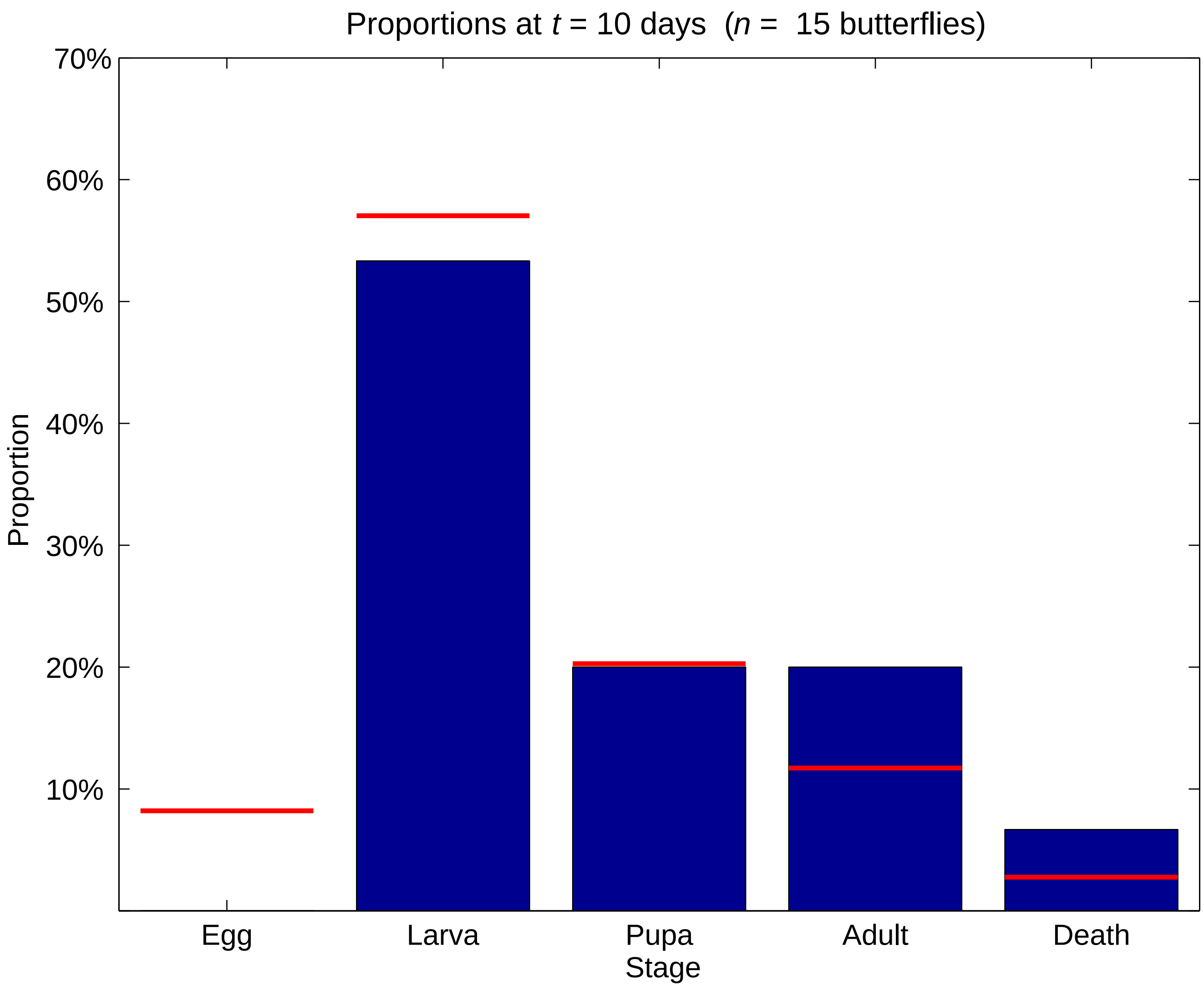
# Ensemble proportions (simulation)



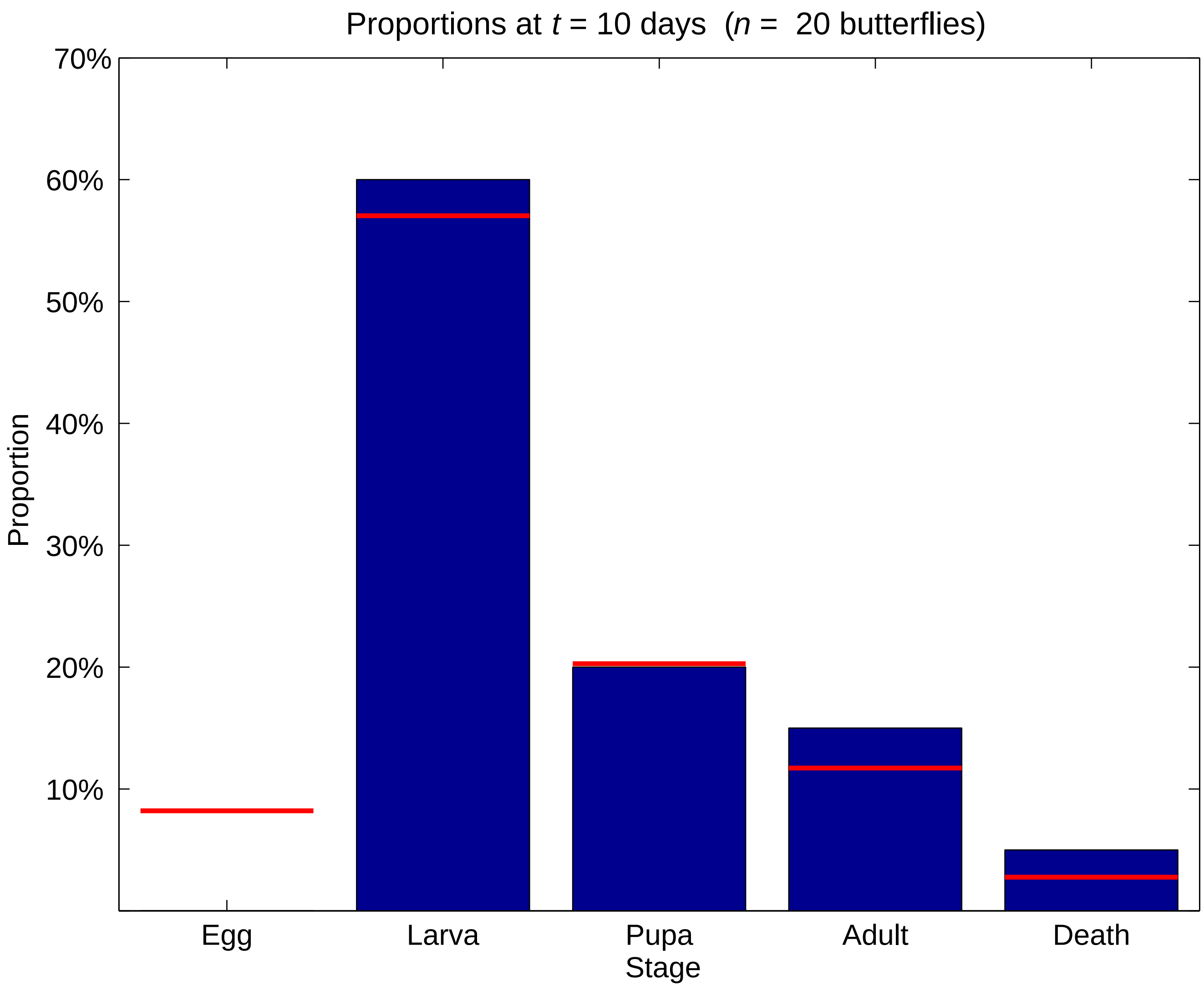
# Ensemble proportions (simulation)



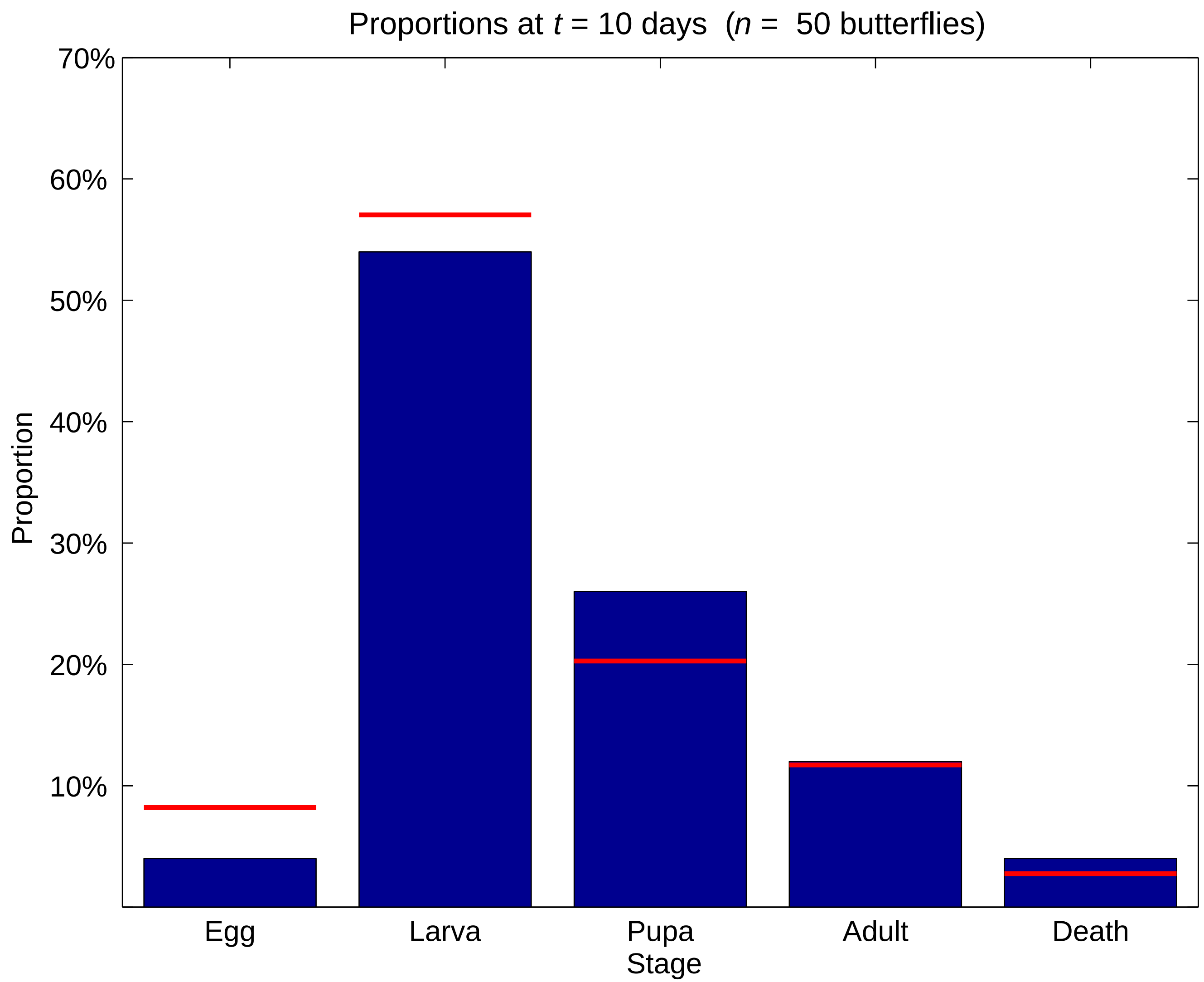
# Ensemble proportions (simulation)



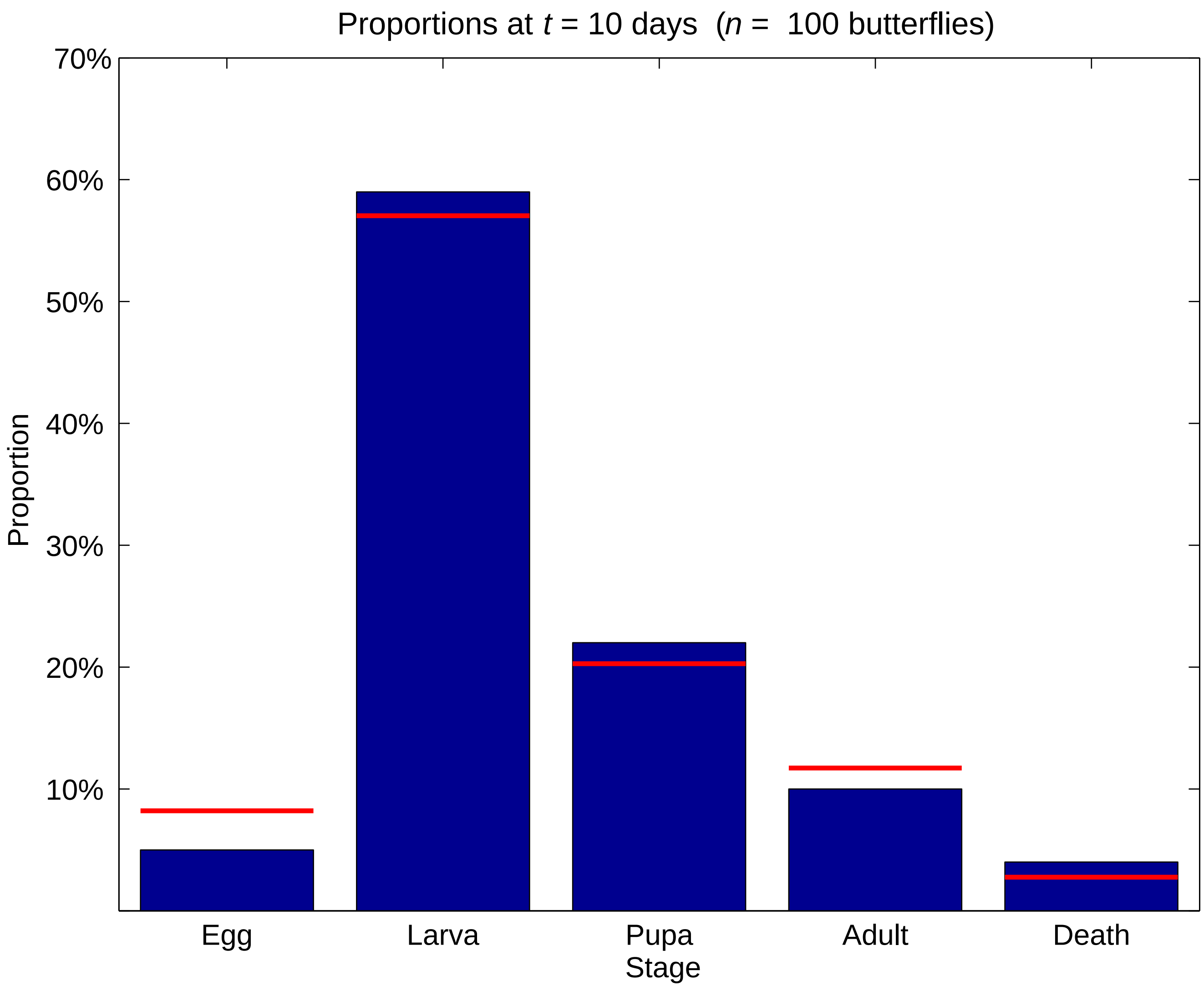
# Ensemble proportions (simulation)



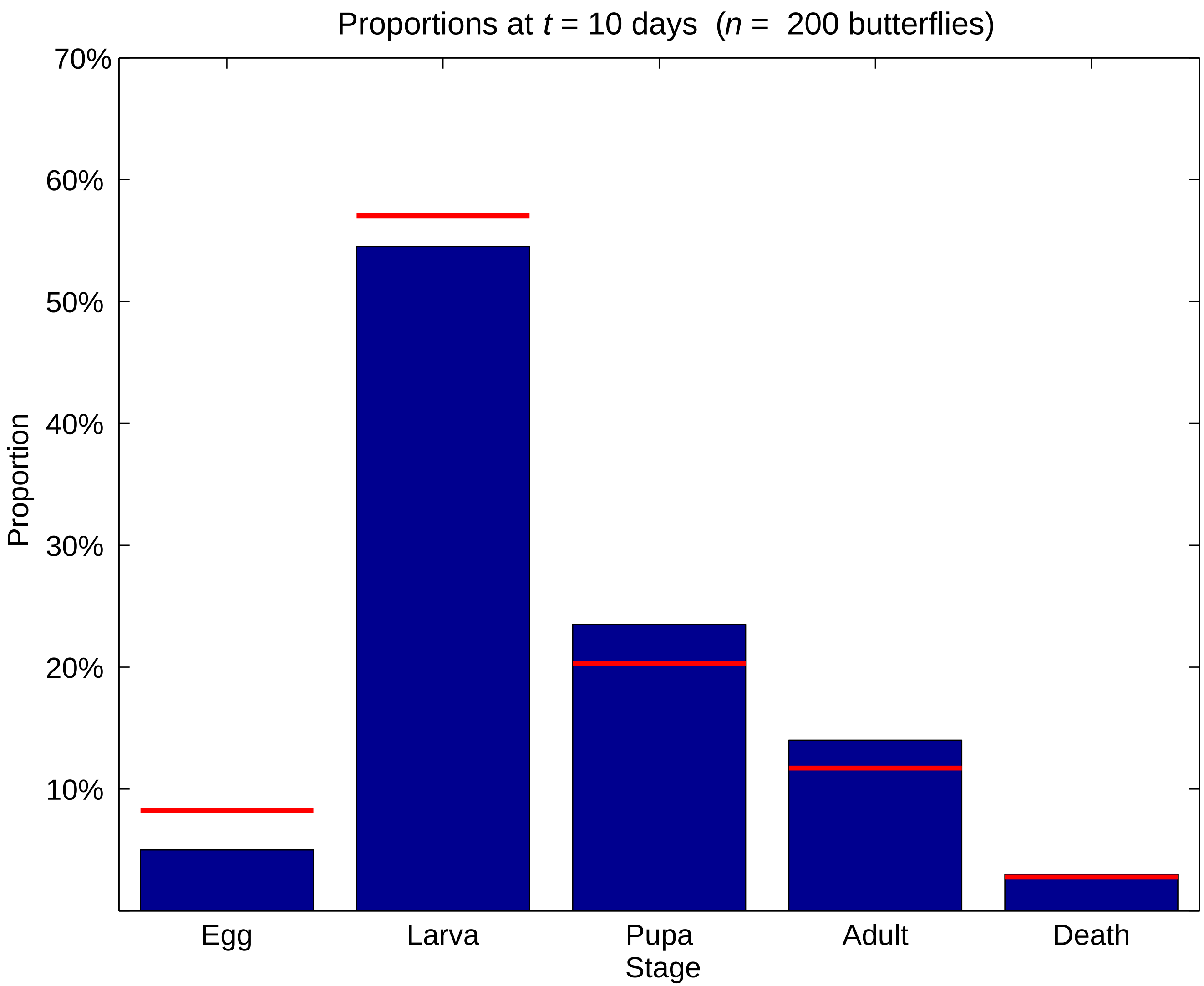
# Ensemble proportions (simulation)



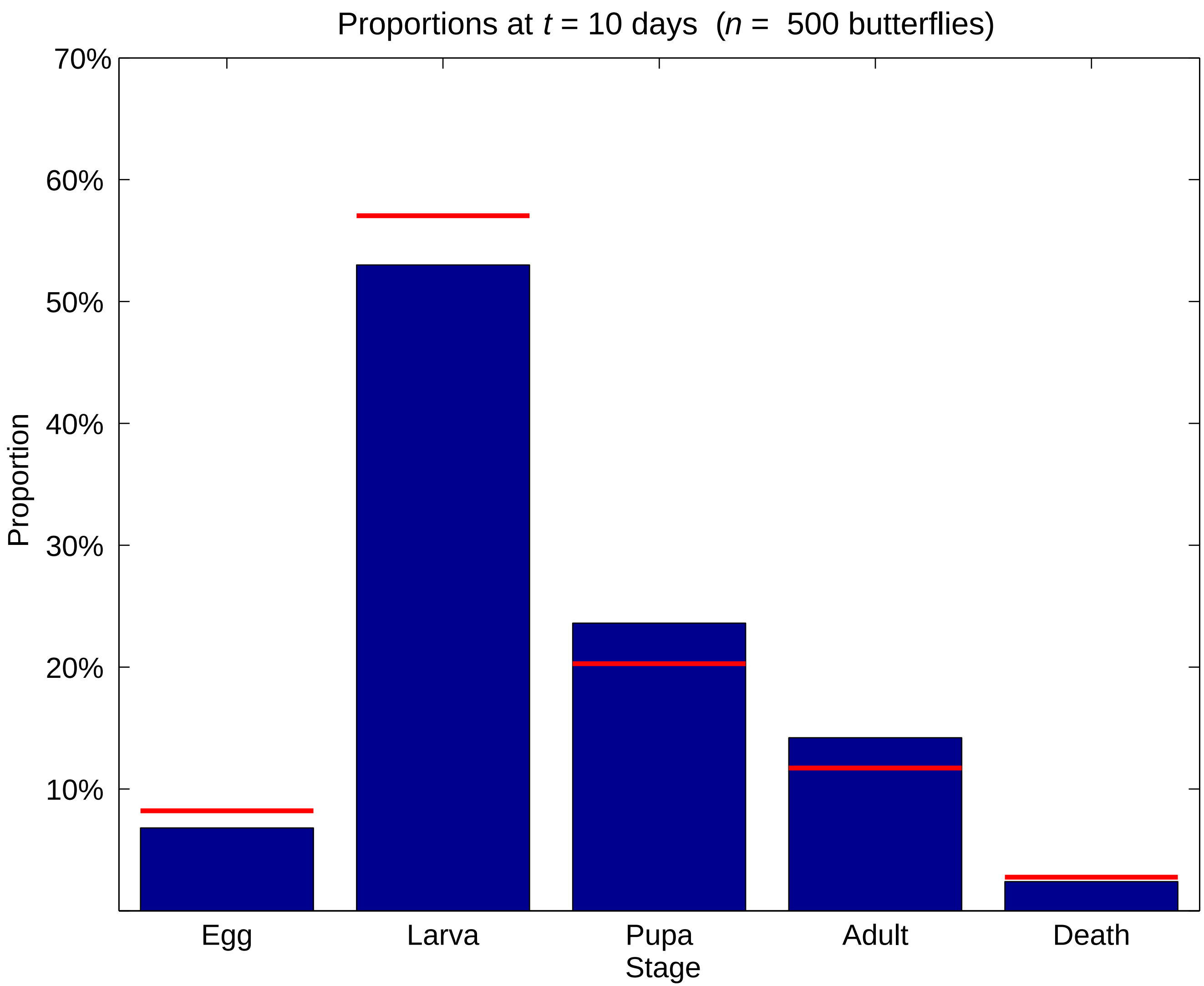
# Ensemble proportions (simulation)



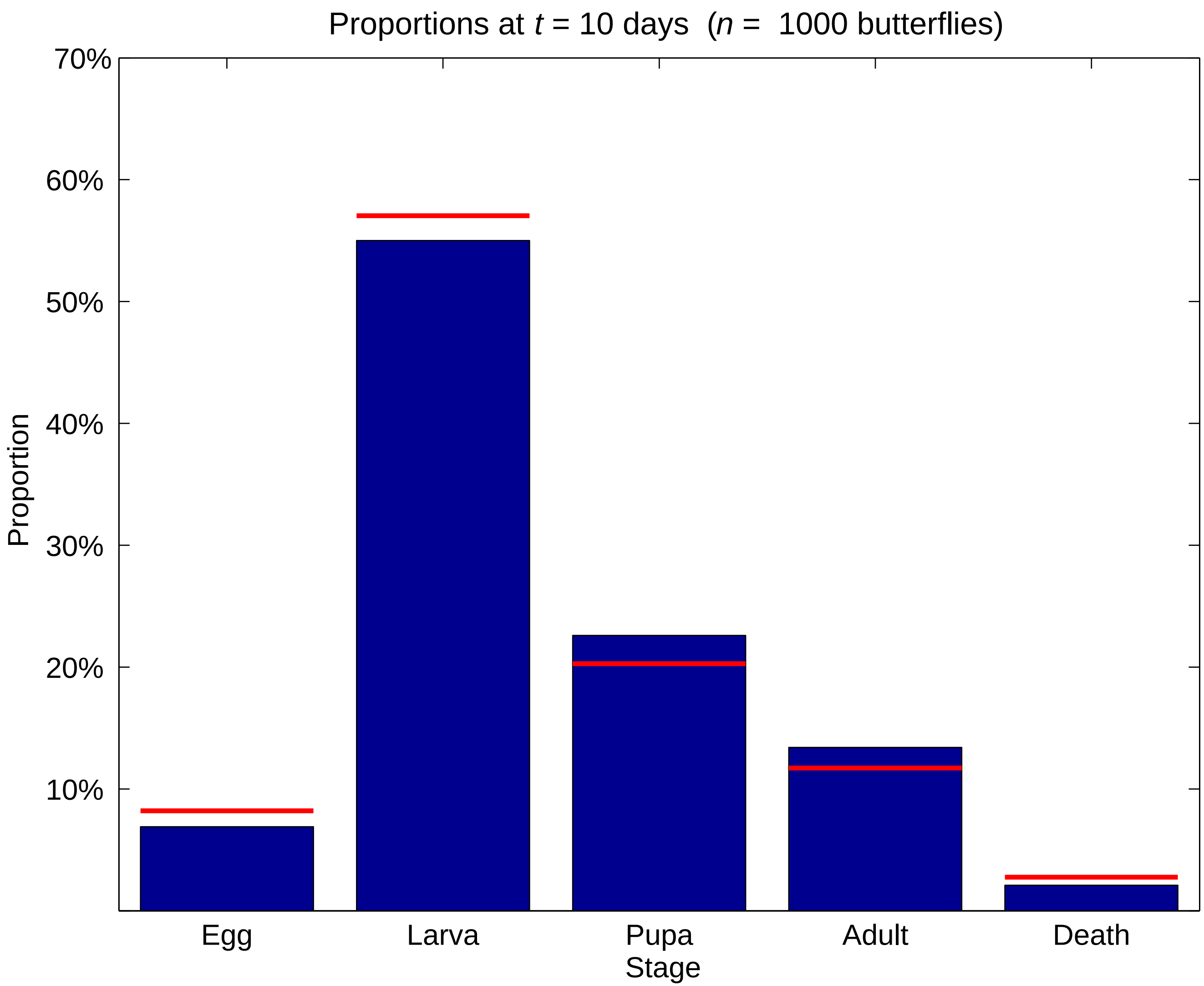
# Ensemble proportions (simulation)



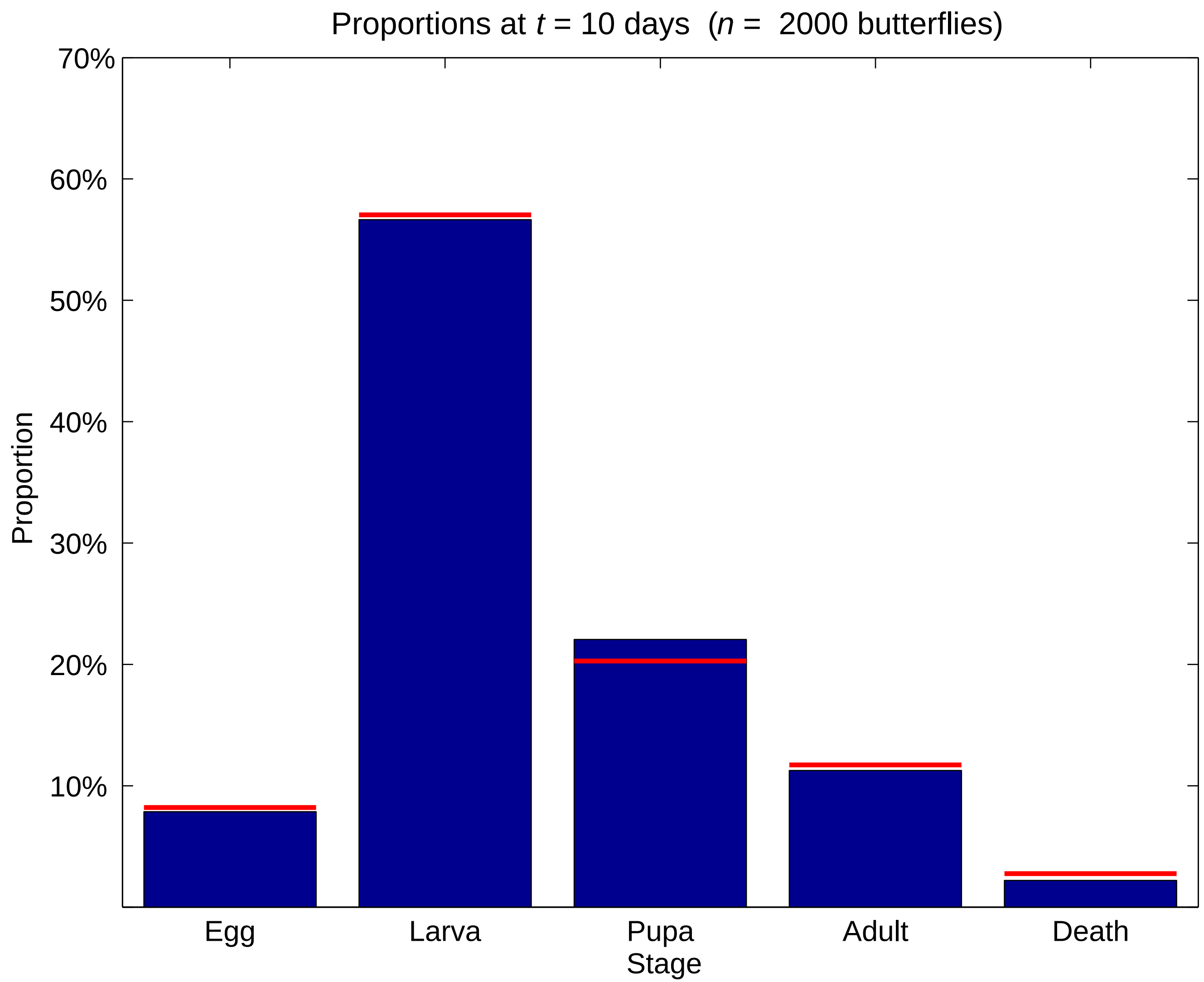
# Ensemble proportions (simulation)



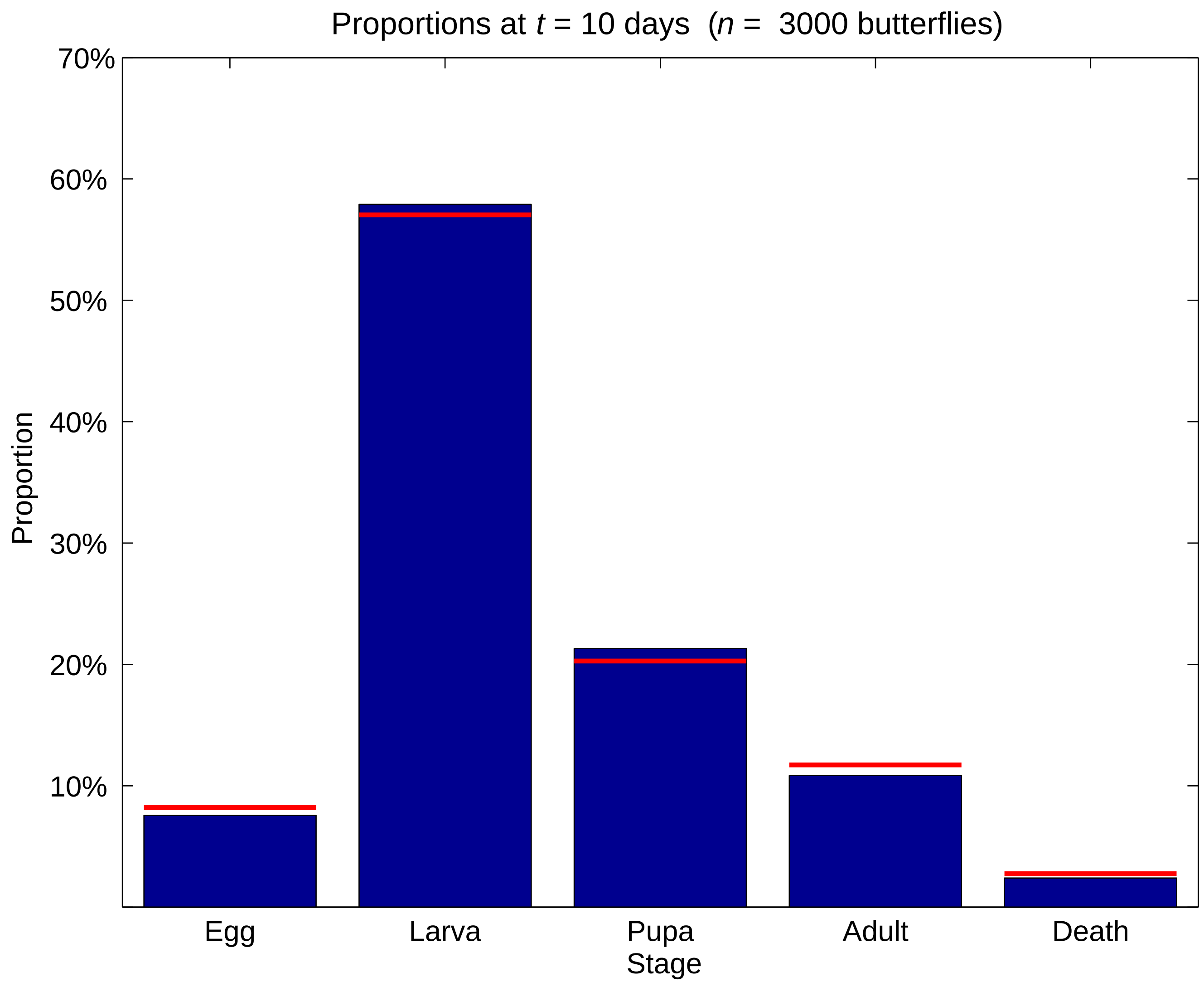
# Ensemble proportions (simulation)



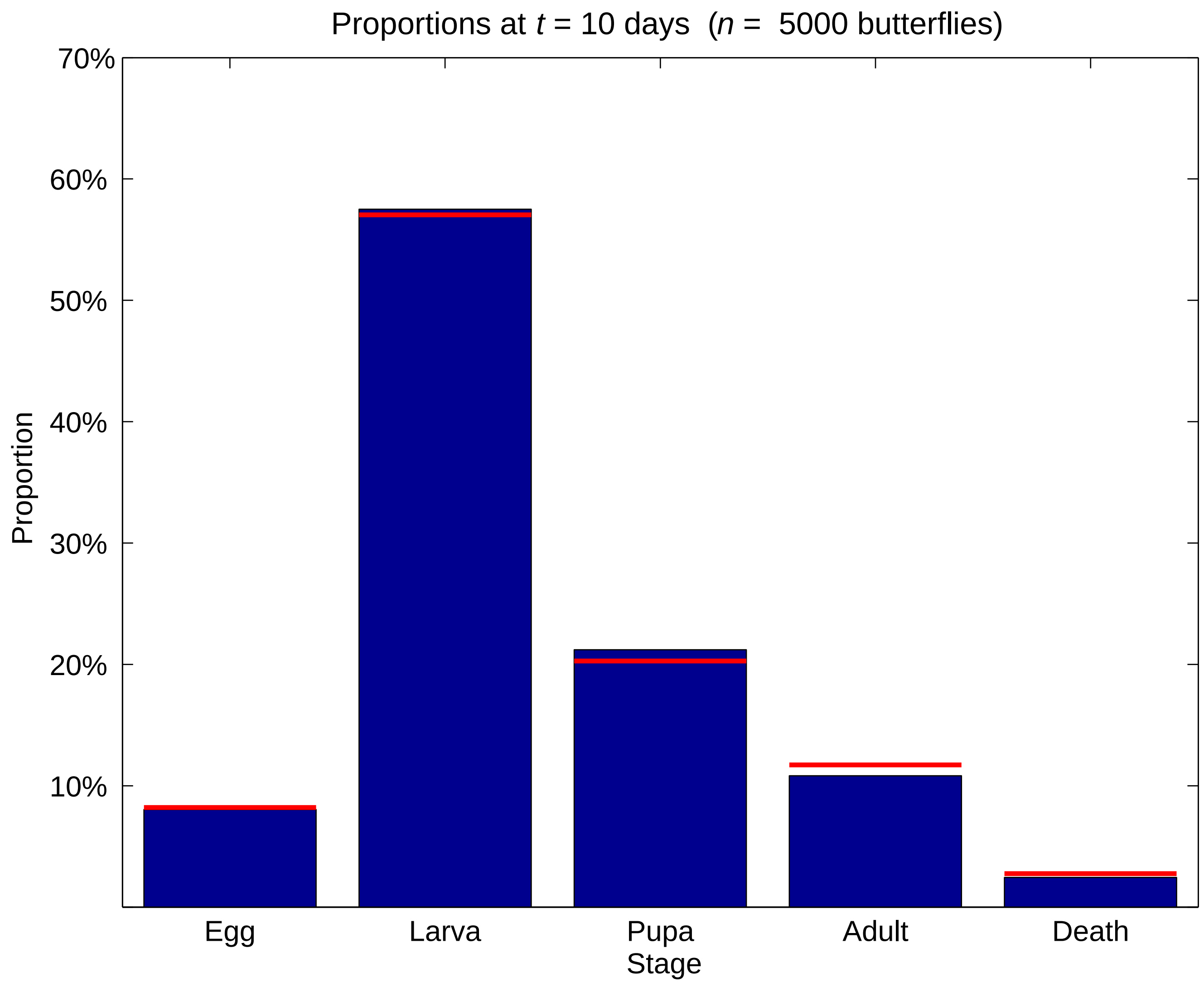
# Ensemble proportions (simulation)



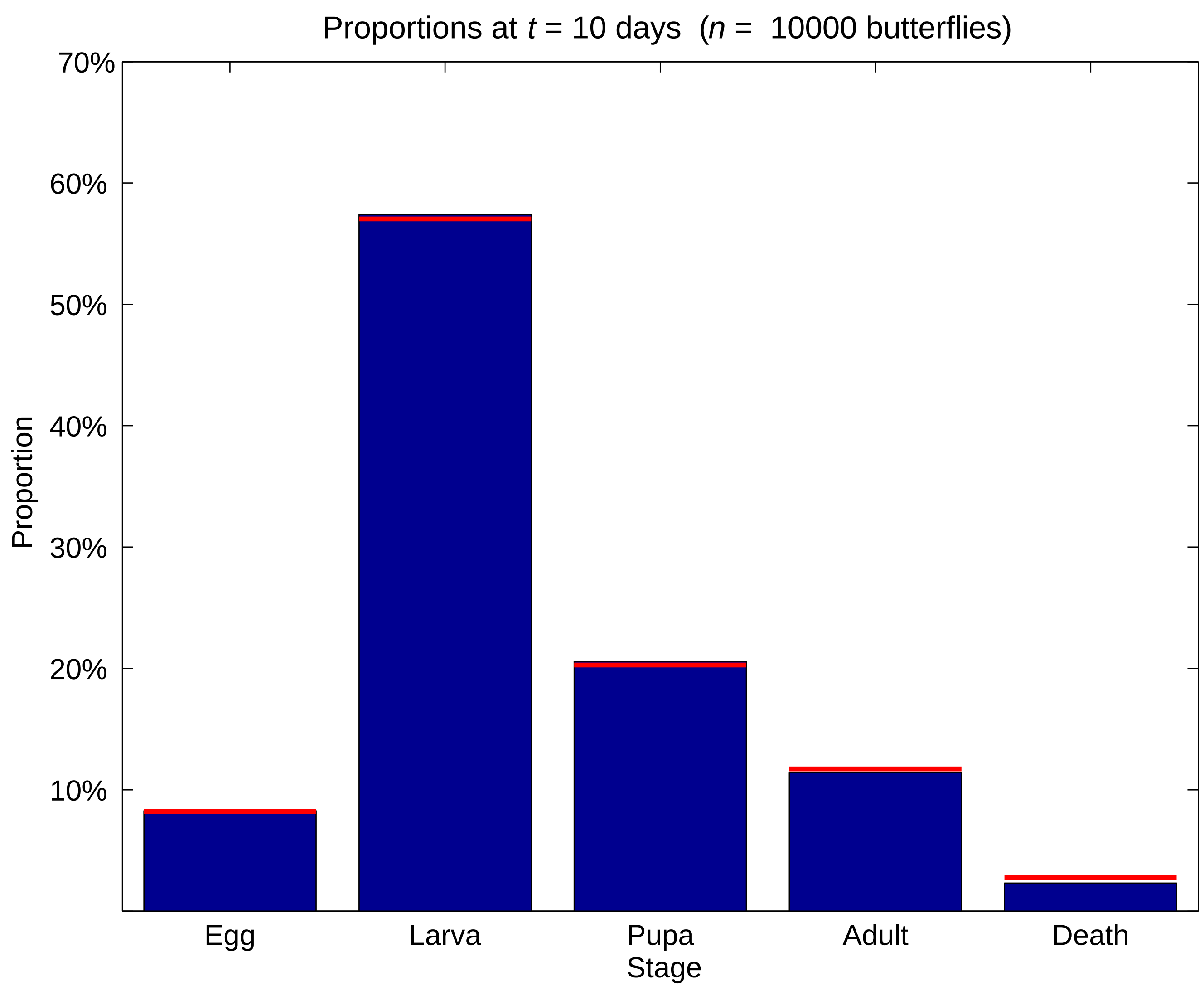
# Ensemble proportions (simulation)



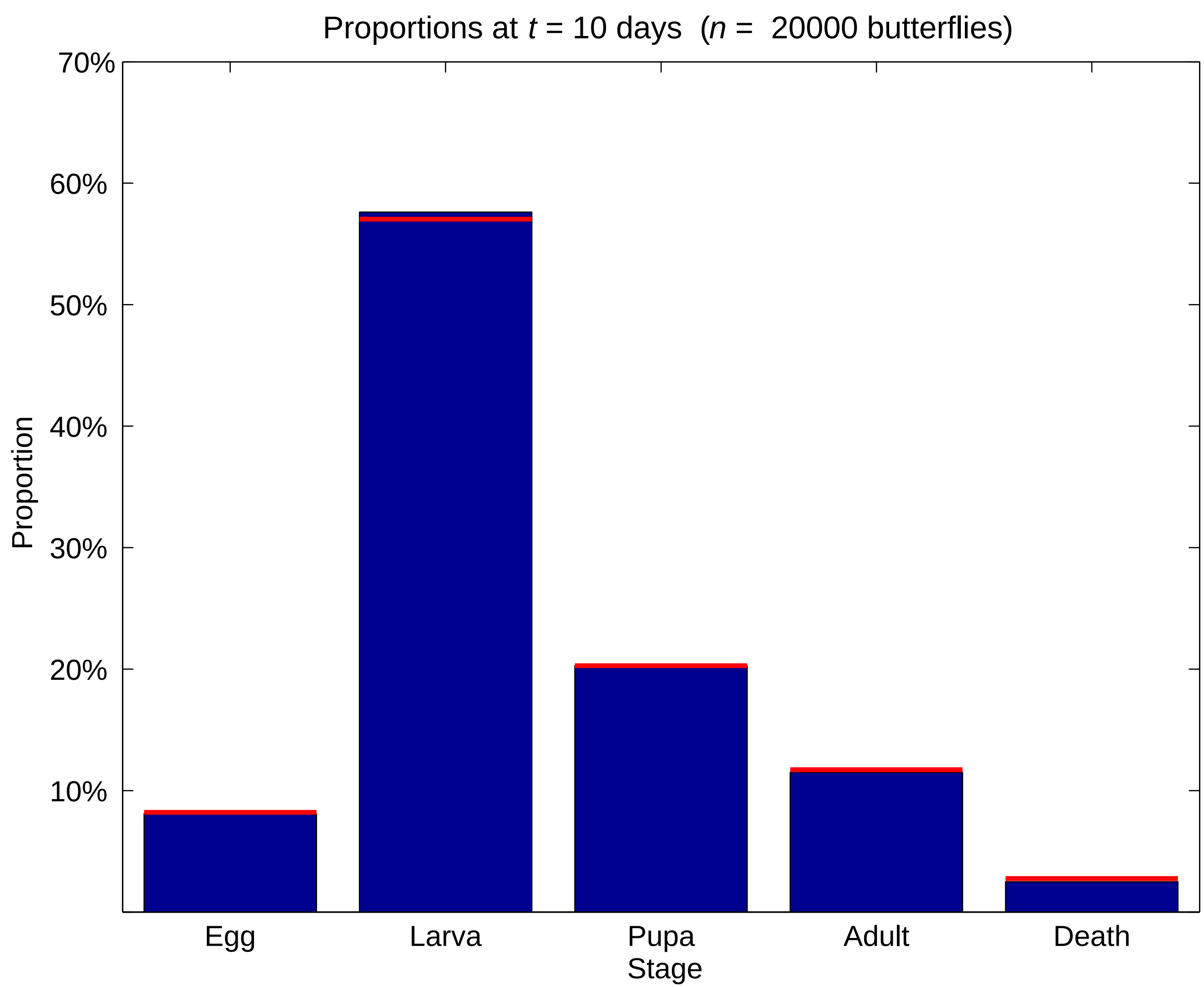
# Ensemble proportions (simulation)



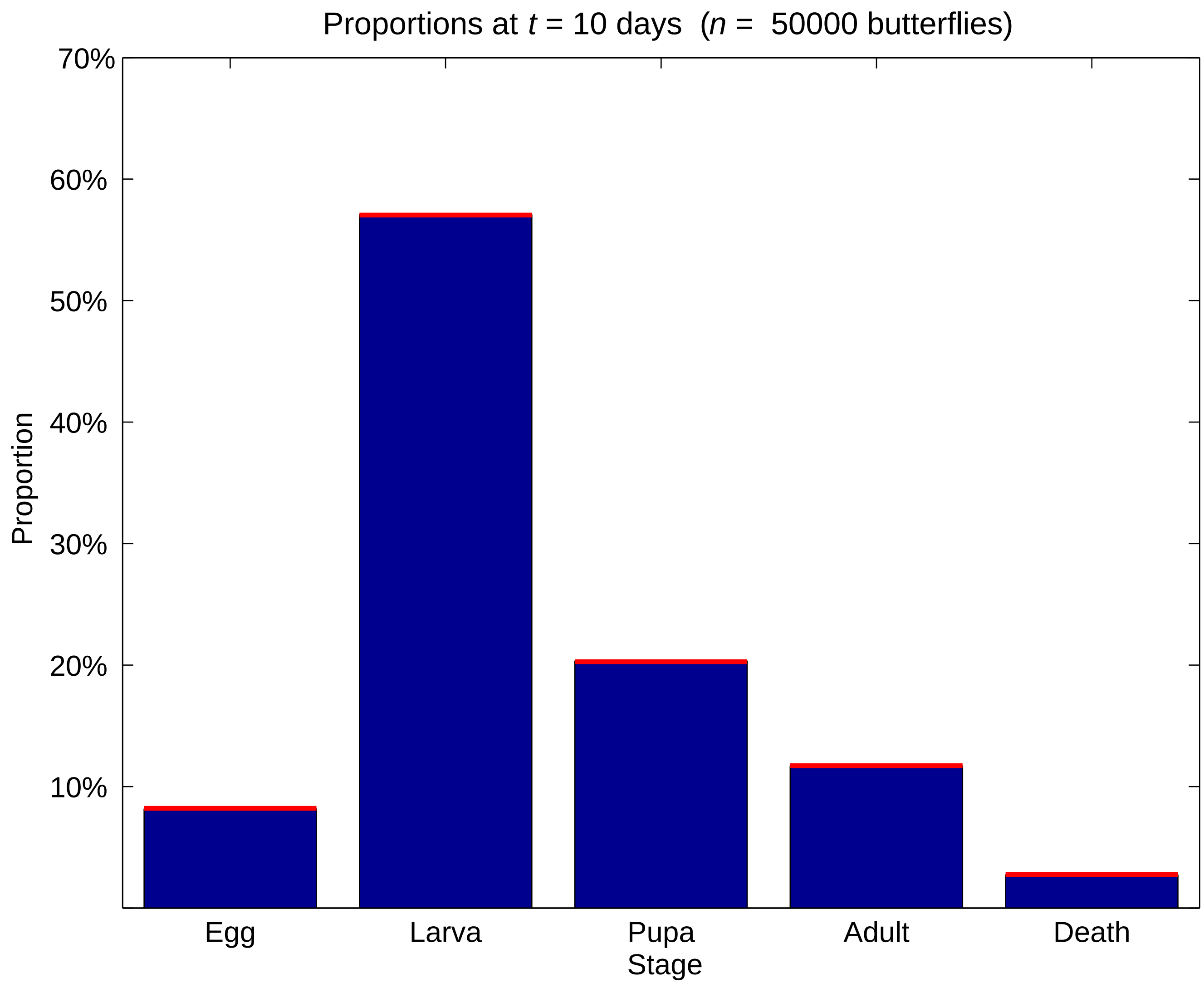
# Ensemble proportions (simulation)



# Ensemble proportions (simulation)



# Ensemble proportions (simulation)



# Convergence of ensemble proportions

Let $\boldsymbol{X}^{(n)}(t) = \boldsymbol{N}(t)/n$, where $n$ is the number of individuals, so that $X_j^{(n)}(t)$ is the proportion if individuals in state $j$.

# Convergence of ensemble proportions

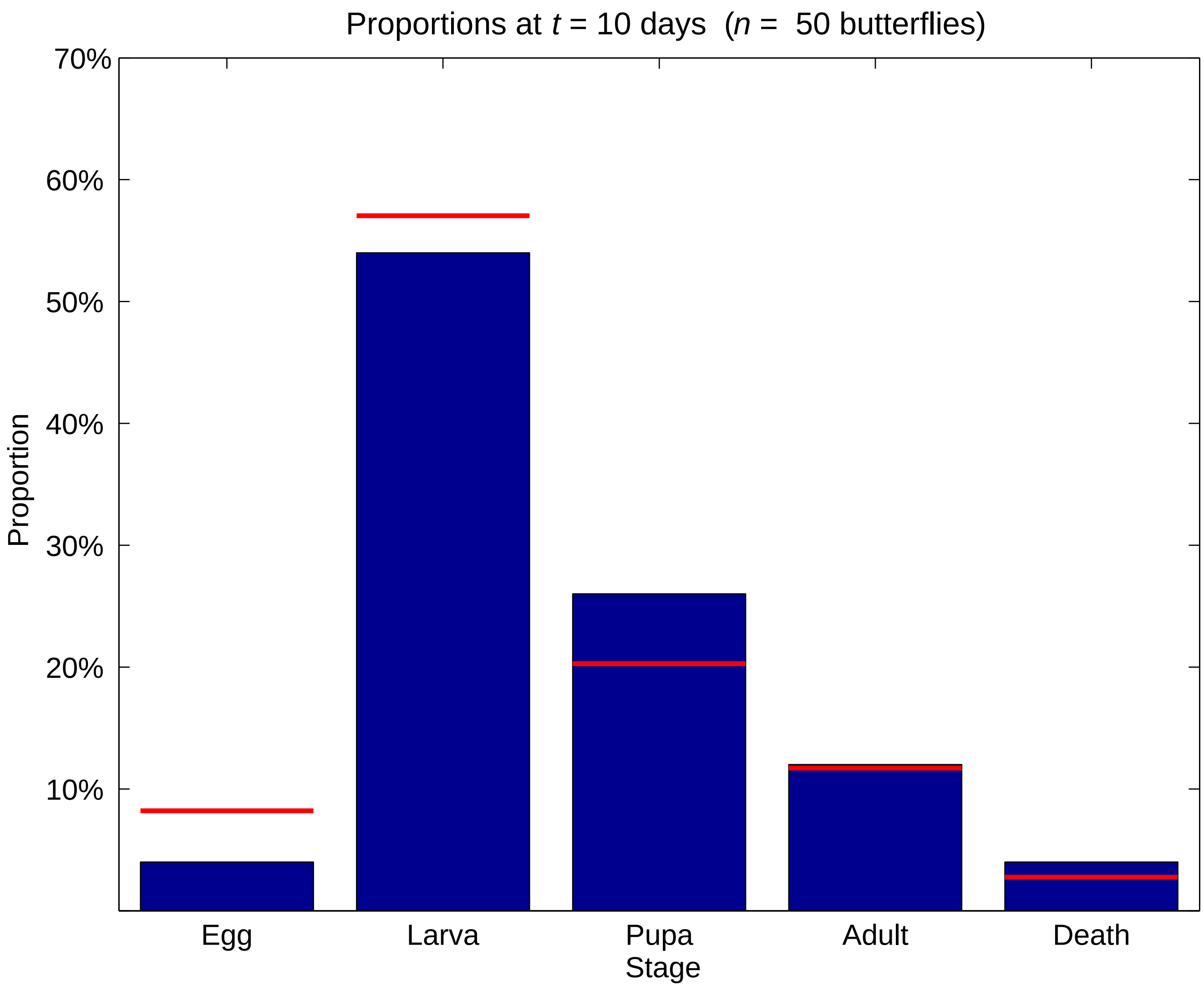
Let $\boldsymbol{X}^{(n)}(t) = \boldsymbol{N}(t)/n$, where $n$ is the number of individuals, so that $X_j^{(n)}(t)$ is the proportion if individuals in state $j$.

**Theorem 1**. If $\boldsymbol{X}^{(n)}(0) \to \boldsymbol{a}$ as $n \to \infty$, then, for all $u > 0$, and for every $\epsilon > 0$,

$$\Pr\left(\sup_{0\le t\le u}\left|\boldsymbol{X}^{(n)}(t) - \boldsymbol{p}(t)\right| > \epsilon\right) \to 0 \quad \text{as } n \to \infty,$$

where $\boldsymbol{p}(t) = (p_j(t), j \in S)$ is the unique solution to $\boldsymbol{p}'(t) = \boldsymbol{p}(t)\,Q$ satisfying $\boldsymbol{p}(0) = \boldsymbol{a}$, namely $\boldsymbol{p}(t) = \boldsymbol{a}\exp(tQ)$, where $\exp(\,\cdot\,)$ is the matrix exponential.

# Bonus theorem



# Bonus theorem

**Theorem 2**. In the setup of Theorem 1, let

$$\boldsymbol{Z}^{(n)}(t) = \sqrt{n}(\boldsymbol{X}^{(n)}(t) - \boldsymbol{p}(t)).$$

If $\boldsymbol{Z}^{(n)}(0) \to \boldsymbol{z}$ as $n \to \infty$, then $(\boldsymbol{Z}^{(n)}(t))$ converges weakly in $D[0,t]$ (the space of right-continuous, left-hand limits functions on $[0,t]$) to a *Gaussian diffusion* $(\boldsymbol{Z}(t))$ with initial value $\boldsymbol{Z}(0) = z$ and with mean and covariance given by $\mu_s := \mathbb{E}(\boldsymbol{Z}(s)) = e^{sQ^\top}\boldsymbol{z}$ and

$$V_s := \mathrm{Cov}(\boldsymbol{Z}(s)) = e^{sQ^\top}\left(\int_0^s e^{-uQ^\top} G(\boldsymbol{p}(u)) e^{-uQ}\, du\right) e^{sQ},$$

# Bonus theorem

**Theorem 2 (continued).**

$\ldots$ where the matrix $G(\boldsymbol{x})$ has entries

$$G_{kk}(\boldsymbol{x}) = x_k q_k + \sum_{i \neq k} x_i q_{ik} \text{ and } G_{kl}(\boldsymbol{x}) = -(x_l q_{lk} + x_k q_{kl}).$$

# Bonus theorem

**Theorem 2 (continued).**

$\ldots$ where the matrix $G(\boldsymbol{x})$ has entries

$$G_{kk}(\boldsymbol{x}) = x_k q_k + \sum_{i \neq k} x_i q_{ik} \text{ and } G_{kl}(\boldsymbol{x}) = -(x_l q_{lk} + x_k q_{kl}).$$

Theorem 2 has many implications. One immediate one is that the population proportions $\boldsymbol{X}^{(n)}(t)$ have an approximate multivariate Gaussian (normal) distribution with known mean vector and covariance matrix.

This helps explain the observed fluctuations (now seen to be of order $1/\sqrt{n}$) of $\boldsymbol{X}^{(n)}(t)$ about $\boldsymbol{p}(t)$.

# Bonus theorem