Here $p(\cdot)$ a prior distribution for parameter vector θ, which may itself have one or more levels of stochastic interdependence. The distribution of each HMM initial state x_{i1} is $f_{i1}(\cdot \mid \theta)$. Markov state transition probabilities are given by $f_{it}(\cdot \mid \theta, x_{i,t-1})$ and observation probabilities by $g_{it}(\cdot \mid \theta, x_{it})$.

Discrete HMMs have long been applied in the area of ecological capture-recapture (*e.g.*, Gimenez et al., 2007; King, 2012; Langrock et al., 2012). In this context, a set of n distinct animals is monitored for k sampling occasions. Each y_i represents the observation history of animal i, for $i = 1, \ldots, n$, which can be modeled using HMMs as in (1). The set of possible observations $\mathcal{Y}$ may include a state to represent "unobserved". Since all n animals are not typically observed on occasion $t = 1$, each embedded HMM will "begin" at the sampling period corresponding to the first genuine observation of that animal.

2.2 Model Likelihood

We now provide the model likelihood for the general HMM formulation in (1), which is used in the Bayesian estimation procedures described next. We begin with the likelihood contribution from a single observation history,

$$L(\theta \mid y_i) = \sum_{x_i \in \mathcal{X}^k} f_{i1}(x_{i1} \mid \theta) \left(\prod_{t=2}^{k} f_{it}(x_{it} \mid \theta, x_{i,t-1}) \right) \left(\prod_{t=1}^{k} g_{it}(y_{it} \mid \theta, x_{it}) \right), \tag{2}$$

where $\mathcal{X}^k$ denotes the standard k-fold Cartesian product of $\mathcal{X}$. Using the likelihood components in (2), the total model likelihood of y is

$$L(\theta \mid y) = \prod_{i=1}^{n} L(\theta \mid y_i).$$

2.3 Computational Approaches

We now describe several computational approaches to applying Bayesian estimation to embedded HMMs. These strategies will form the basis for our comparisons, using examples