Ecological Archives A/E/M000-000-A1

Jeff A. Tracey, Jun Zhu, Erin Boydston, Lisa Lyren, Robert N. Fisher, and Kevin R. Crooks. 2012. Mapping behavioral landscapes for animal movement: A finite mixture modeling approach. *Ecological Applications* **VOL: pp-pp**

Appendix B: Likelihood-Based Inference Using Particle Swarm Optimization and EM Algorithm.

B.1 Optimization of the negative log-likelihood functions

In this section, we describe methods for obtaining maximum likelihood estimates of the parameters in the finite mixture models for move angles, move distances, and both move angles and distances. Finite mixture models can be difficult to optimize because they can contain numerous local optima (McLachlan and Peel, 2000). Furthermore, most numerical optimization routines are prone to becoming trapped by these local optima (Hassoun, 1995). This is exacerbated by the fact that the optimum to which the algorithm converges is strongly dependent on the initial parameter values supplied to the algorithm. Good initial parameter values can be difficult to select, particularly for models with many parameters. To remedy this situation, we repeat the overall optimization procedure multiple times and retain the result with the lowest negative log-likelihood agreed upon for a given model. The optimization procedure that we use consists of two steps: (1) particle swarm optimization to find initial parameter values for numerical optimization and (2) numerical optimization using an Expectation - Maximization algorithm.

Particle swarm optimization (PSO) is a stochastic optimization algorithm (Poli et al., 2007). An advantage of PSO is that it is more resistant to being trapped by local optima than most numerical optimization algorithms. The algorithm is motivated by natural systems that demonstrate swarm intelligence, such as flocking birds and swarming insects. In PSO, a population of particles move through parameter space and share information with each other while doing so. The movement of each particle is updated by adding a vector giving the location of the particle in the previous iteration and a vector of velocities. Each particle's velocity is determined by a stochastically weighted sum of its velocity at the previous iteration, the difference between the particle's best solution and its position at the previous iteration, and the difference between the best solution found by all neighboring particles and its position at the previous iteration. Numerous variations on the basic algorithm exist. After PSO is performed for a predetermined number of iterations, the best solution found by any particle is passed to the Expectation - Maximization algorithm as the initial value for the parameters.

Following PSO, numerical optimization is performed using an Expectation - Maximization (EM) algorithm (Dempster et al., 1977). It is possible to directly optimize the observed (incomplete) data negative log-likelihood function, but from initial tests that we conducted we concluded that use of the EM algorithm leads to better convergence. The main idea behind the EM algorithm is to formulate the optimization problem as one of incomplete data; that is, that we treat the component label Z_g that identifies the mixture component

that generated the ith observed response as missing data (McLachlan and Peel, 2000). The algorithm consists of an E-step and an M-step. In the E-step, the expected probabilities that $Z_g = z_g$ given the current estimates of the parameters are computed. In the M-step, the value of parameters that maximize the expectation of the complete-data log-likelihood (ℓ_c) is found by numerical optimization. The E-step and M-step are iterated until convergence. An example in which this approach is applied to simulated data is given in Appendix C.

B.2 EM Algorithm

B.2.1 Move angles

Let $\mathbf{Y} = (Y_1, \ldots, Y_I)'$ denote the vector of observed move angles. Also let $\mathbf{R} = (R_1, \ldots, R_I)'$ denote the vector of observed response angles in relation to the angular covariates $\mathbf{A} = (A_1, \ldots, A_I)'$ and $\mathbf{X}_i = (\mathbf{X}'_{i,0}, \ldots, \mathbf{X}'_{i,J})'$ denote the vector of covariates related to the selection of a movement response by the animal at location $\mathbf{s}_i$. Recall that the component mixing proportion parameters are $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_G)'$ with $\boldsymbol{\beta}_1 = \mathbf{0}$ and $\boldsymbol{\beta}_g = (\beta_{g,0}, \ldots, \beta_{g,J})'$ when $g > 1$. Also recall that the component pdf parameters are $\boldsymbol{\eta} = (\boldsymbol{\eta}'_1, \ldots, \boldsymbol{\eta}'_G)'$, where $\boldsymbol{\eta}_g = (\boldsymbol{\mu}'_g, \boldsymbol{\kappa}'_g, \boldsymbol{\phi}'_g)'$, $\boldsymbol{\mu}_g = (\mu_{g,1}, \ldots, \mu_{g,s_g})'$, $\boldsymbol{\kappa}_g = (\kappa_{g,1}, \ldots, \kappa_{g,s_g})'$, and $\boldsymbol{\phi}_g = (\phi_{g,1}, \ldots, \phi_{g,s_g})'$. In addition, let $\boldsymbol{\theta}_Y = (\boldsymbol{\beta}', \boldsymbol{\eta}')'$. Assuming conditional independence of the move angles conditional on $\{\mathbf{X}_i\}$ and $\{A_i\}$ the log-likelihood function for $\boldsymbol{\theta}_Y$ is:

$$\begin{aligned} \ell(\boldsymbol{\theta}_Y; \mathbf{Y}, \mathbf{A}, \mathbf{X}) &= \sum_{i=1}^{I} \log \{f(Y_i|\mathbf{X}_i, A_i; \boldsymbol{\theta}_Y)\} \\ &= \sum_{i=1}^{I} \log \left\{ \sum_{g=1}^{G} p_g(\mathbf{X}_i; \boldsymbol{\beta}) f_g(Y_i|A_i; \boldsymbol{\eta}_g) \right\}. \end{aligned} \tag{B.1}$$

Maximization of the log-likelihood function (B.1) can be obtained using an EM algorithm (McLachlan and Peel, 2000) as follows. The observed move angles $\mathbf{Y}$ are associated with component-labels $\mathbf{U} = (\mathbf{U}'_1, \ldots, \mathbf{U}'_I)$, where $\mathbf{U}_i = (U_{i1}, \ldots, U_{iG})'$, $U_{ig} = \mathcal{I}(Y_i$ arises from the gth component). Suppose

$$\mathbf{U}_i \sim \text{Multinomial}_G(1, \mathbf{p}_i),$$

where $\mathbf{p}_i = (p_{i1}, \ldots, p_{iG})' = (p_1(\mathbf{X}_i, \boldsymbol{\beta}), \ldots, p_G(\mathbf{X}_i, \boldsymbol{\beta}))'$ and $\{\mathbf{U}_i : i = 1, \ldots, I\}$ are independent. Then the complete-data log-likelihood for $\boldsymbol{\theta}_Y$ is

$$\ell_c(\boldsymbol{\theta}_Y; \mathbf{Y}, \mathbf{A}, \mathbf{U}) = \sum_{g=1}^{G} \sum_{i=1}^{I} U_{ig} \left\{ \log p_g(\mathbf{X}_i; \boldsymbol{\beta}) + \log f_g(Y_i|A_i; \boldsymbol{\eta}_g) \right\}.$$

In the E-step, we compute the expectation of ℓ_c conditional on $\mathbf{Y}$ and evaluated at $\hat{\boldsymbol{\theta}}_Y$:

$$\begin{aligned} Q_Y(\boldsymbol{\theta}_Y; \hat{\boldsymbol{\theta}}_Y) &= E\left\{ \ell_c(\boldsymbol{\theta}_Y; \mathbf{Y}, \mathbf{A}, \mathbf{X}, \mathbf{U}) | \mathbf{Y}, \mathbf{A}, \mathbf{X}; \hat{\boldsymbol{\theta}}_A \right\} \\ &= \sum_{g=1}^{G} \sum_{i=1}^{I} \hat{U}_{ig} \left\{ \log p_{ig} + \log f_g(Y_i|A_i; \boldsymbol{\eta}_g) \right\}, \end{aligned}$$

where $\hat{U}_{ig} = E\left(U_{ig}|\mathbf{Y};\hat{\boldsymbol{\theta}}_Y\right)$. It is straightforward to show that,

$$\hat{U}_{ig} = \hat{p}_{ig} f_g(Y_i;\hat{\boldsymbol{\eta}}_g)\left\{\sum_{g=1}^{G}\hat{p}_{ig} f_g(Y_i;\hat{\boldsymbol{\eta}}_g)\right\}^{-1}; \quad g = 1,\ldots,G.$$

In the M-step, we maximize $Q_Y(\boldsymbol{\theta}_Y;\hat{\boldsymbol{\theta}}_Y)$ with respect to $\boldsymbol{\theta}_Y$.

B.2.2 Move distances

Let $\mathbf{D} = (D_1,\ldots,D_I)'$ denote the vector of observed move distances. Also let $\mathbf{X}_i = (\mathbf{X}'_{i,0}$,..., $\mathbf{X}'_{i,J})'$ denote the vector of covariates related to the selection of a movement response by the animal at location $\mathbf{s}_i$. As above, $\boldsymbol{\beta}$ is a vector of mixing proportion parameters. Also recall that the component pdf parameters are $\boldsymbol{\zeta} = (\boldsymbol{\zeta}'_1,\ldots,\boldsymbol{\zeta}'_G)'$, where $\boldsymbol{\zeta}_g = (\boldsymbol{\alpha}'_g,\boldsymbol{\lambda}'_g,\boldsymbol{\phi}'_g)'$, $\boldsymbol{\alpha}_g = (\alpha_{g,1},\ldots,\alpha_{g,s_g})'$, $\boldsymbol{\lambda}_g = (\lambda_{g,1},\ldots,\lambda_{g,s_g})'$, and $\boldsymbol{\phi}_g = (\phi_{g,1},\ldots,\phi_{g,s_g})'$. In addition, let $\boldsymbol{\theta}_D = (\boldsymbol{\beta}',\boldsymbol{\zeta}')'$ denote the parameters of the move distance model. Assuming conditional independence of the move distances conditional on $\{\mathbf{X}_i\}$, the log-likelihood function for $\boldsymbol{\theta}_D$ is:

$$\begin{aligned}\ell(\boldsymbol{\theta}_D;\mathbf{D},\mathbf{X}) &= \sum_{i=1}^{I}\log\{f(D_i|\mathbf{X}_i;\boldsymbol{\theta}_D)\}\\ &= \sum_{i=1}^{I}\log\left\{\sum_{g=1}^{G}p_g(\mathbf{X}_i;\boldsymbol{\beta})h_g(D_i|\boldsymbol{\zeta}_g)\right\}. \qquad \text{(B.2)}\end{aligned}$$

Maximization of the log-likelihood function (B.2) again can be obtained using an EM algorithm. The observed move distances $\mathbf{D}$ are associated with component-labels $\mathbf{V} = (\mathbf{V}'_1,\ldots,\mathbf{V}'_I)$, as specified in for the move angle model. Then the complete-data log-likelihood for $\boldsymbol{\theta}_D$ is

$$\ell_c(\boldsymbol{\theta}_D;\mathbf{D},\mathbf{V}) = \sum_{g=1}^{G}\sum_{i=1}^{I}V_{ig}\left\{\log p_g(\mathbf{X}_i;\boldsymbol{\beta}) + \log h_g(D_i|\boldsymbol{\zeta}_g)\right\}.$$

In the E-step, we compute the expectation of ℓ_c conditional on $\mathbf{D}$ and evaluated at $\hat{\boldsymbol{\theta}}_D$:

$$\begin{aligned}Q_D(\boldsymbol{\theta}_D;\hat{\boldsymbol{\theta}}_D) &= E\left\{\ell_c(\boldsymbol{\theta}_D;\mathbf{D},\mathbf{V})|\mathbf{D};\hat{\boldsymbol{\theta}}_D\right\}\\ &= \sum_{g=1}^{G}\sum_{i=1}^{I}\hat{V}_{ig}\left\{\log p_g(\mathbf{X}_i;\hat{\boldsymbol{\beta}}_g) + \log h_g(D_i|\hat{\boldsymbol{\zeta}}_g)\right\},\end{aligned}$$

where $\hat{V}_{ig} = E(V_{ig}|\mathbf{D};\hat{\boldsymbol{\theta}}_D)$. As in the move angle model, we have,

$$\hat{V}_{ig} = \hat{p}_{ig}h_g(D_i;\hat{\boldsymbol{\zeta}}_g)\left\{\sum_{g=1}^{G}\hat{p}_{ig}h_g(D_i;\hat{\boldsymbol{\zeta}}_g)\right\}^{-1}; \quad g = 1,\ldots,G,$$

where $\hat{p}_{ig} = p_g(\mathbf{X}_i;\hat{\boldsymbol{\beta}})$ for $g = 1,\ldots,G$. In the M-step, we maximize $Q_D(\boldsymbol{\theta}_D;\ \hat{\boldsymbol{\theta}}_D)$ with respect to $\boldsymbol{\theta}_D$.

B.2.3 Move angles and distances

Here we describe inference when modeling both move angles and distances. The parameters follow the notation for the move angle and move distance models above. Here, we let $\boldsymbol{\theta}_{YD} = (\boldsymbol{\beta}', \boldsymbol{\eta}', \boldsymbol{\zeta}')'$ denote the parameters of the combined move angle and distance model described in (**??**). Assuming conditional independence of the move angles and distances conditional on $\{\mathbf{X}_i\}$, the log-likelihood function for $\boldsymbol{\theta}_{YD}$ is:

$$\begin{aligned}
&\ell(\boldsymbol{\theta}_{YD}; \mathbf{Y}, \mathbf{D}, \mathbf{A}, \mathbf{X}) \\
&= \sum_{i=1}^{I} \log \{f(Y_i, D_i | A_i, \mathbf{X}_i; \boldsymbol{\theta}_{YD})\} \\
&= \sum_{i=1}^{I} \log \left\{ \sum_{g=1}^{G} p_g(\mathbf{X}_i; \boldsymbol{\beta}) f_g(Y_i | A_i; \boldsymbol{\eta}_g) h_g(D_i | \boldsymbol{\zeta}_g) \right\}. \qquad \text{(B.3)}
\end{aligned}$$

As before, maximization of the log-likelihood function (B.3) again can be processed using the EM algorithm. The observed move distances $\mathbf{D}$ are associated with component-labels $\mathbf{W} = (\mathbf{W}_1', \ldots, \mathbf{W}_I')$, as specified in for the move angle models. Then the complete-data log-likelihood for $\boldsymbol{\theta}_{YD}$ is

$$\begin{aligned}
&\ell_c(\boldsymbol{\theta}_{YD}; \mathbf{Y}, \mathbf{D}, \mathbf{A}, \mathbf{W}) \\
&= \sum_{g=1}^{G} \sum_{i=1}^{I} W_{ig} \left\{ \log p_g(\mathbf{X}_i; \boldsymbol{\beta}) + \log f_g(Y_i | A_i; \boldsymbol{\eta}_g) + \log h_g(D_i | \boldsymbol{\zeta}_g) \right\}.
\end{aligned}$$

In the E-step, we compute the expectation of ℓ_c conditional on the observed data and evaluated at $\hat{\boldsymbol{\theta}}_{YD}$:

$$\begin{aligned}
&Q_{YD}(\boldsymbol{\theta}_{YD}; \hat{\boldsymbol{\theta}}_{YD}) = E\left\{ \ell_c(\boldsymbol{\theta}_{YD} | \mathbf{Y}, \mathbf{D}, \mathbf{A}, \mathbf{W}); \hat{\boldsymbol{\theta}}_{YD} \right\} \\
&= \sum_{g=1}^{G} \sum_{i=1}^{I} \hat{W}_{ig} \left\{ \log p_g(\mathbf{X}_i; \hat{\boldsymbol{\beta}}) + \log f_g(Y_i | A_i; \boldsymbol{\eta}_g) + \log h_g(D_i | \hat{\boldsymbol{\zeta}}_g) \right\},
\end{aligned}$$

where $\hat{W}_{ig} = E(W_{ig} | \mathbf{D}; \hat{\boldsymbol{\theta}}_D)$. As in the move angle model, we have,

$$\hat{W}_{ig} = \hat{p}_{ig} f_g(Y_i | A_i; \hat{\boldsymbol{\eta}}_g) h_g(D_i | \hat{\boldsymbol{\zeta}}_g) \left\{ \sum_{g=1}^{G} \hat{p}_{ig} f_g(Y_i | A_i; \hat{\boldsymbol{\eta}}_g) h_g(D_i | \hat{\boldsymbol{\zeta}}_g) \right\}^{-1}$$

where $\hat{p}_{ig} = p_g(\mathbf{X}_i; \hat{\boldsymbol{\beta}})$ for $g = 1, \ldots, G$. In the M-step, we maximize $Q_{YD}(\boldsymbol{\theta}_{YD}; \hat{\boldsymbol{\theta}}_{YD})$ with respect to $\boldsymbol{\theta}_{YD}$.

References

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society B **39**:1–38.

Hassoun, M. H. 1995. Fundamentals of Artificial Neural Networks. MIT Press.

McLachlan, P., and D. Peel. 2000. Finite Mixture Models. Wiley.

Poli, R., J. Kennedy, and T. Blackwell. 2007. Particle swarm optimization: an overview. Swarm Intelligence **1**:33–57.