Haul-out Site Estimation using a Truncation Approximation of a Dirichlet Process Mixture Model

Brian M. Brost

21 July 2015

Implementation

The file haulout.dp.mixture.mu_t.sim.R simulates data according to the model statement presented below, and haulout.dp.mixture.mu_t.mcmc.R contains the MCMC algorithm for model fitting. Model implementation follows the blocked Gibbs sampler truncation approximation of Ishwaran and James (2001) and Gelman et al. (2014).

Model statement

Let $\mathbf{s}_t = (s_{1,t}, s_{2,t})'$ and $\boldsymbol{\mu}_t = (\mu_{1,t}, \mu_{2,t})'$, for $t = 1, \ldots, T$, be observed and true locations, respectively. Also let $\boldsymbol{\mu}_{0,h} = (\mu_{0,1,h}, \mu_{0,2,h})'$, for $h = 1, \ldots, H$, be the locations of haul-out sites (i.e., clusters), where H is the maximum number of haul-outs allowed per the truncation approximation of the Dirichlet process mixture model. Let z_t be a latent indicator variable that denotes when locations are on a haul-out site ($z_t = 1$) or not ($z_t = 0$). Note that when an individual is hauled-out (i.e, $z_t = 1$), $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{0,h_t}$, where h_t acts as a classification variable that identifies the $\boldsymbol{\mu}_{0,h}$ associated with each $\boldsymbol{\mu}_t$. Furthermore, let $\mathcal{S}$ be the support of the movement process and $\tilde{S}$ be the support of the haul-out sites (i.e., the Dirichlet process and all possible $\boldsymbol{\mu}_{0,h}$). Note that $\mathcal{S}$ and $\tilde{S}$ overlap, i.e., $\tilde{\mathcal{S}} \subset \mathcal{S}$. The domain defined by $\mathcal{S}$ therefore represents at-sea locations or locations of the individual while milling adjacent to the haul-out site. Also note that $\bar{\mathcal{S}}$, the complement of $\mathcal{S}$, represents inaccessible locations (i.e., terrestrial sites that are not haul-outs).

$$
\begin{aligned}
\mathbf{s}_t &\sim \mathcal{N}(\boldsymbol{\mu}_t, \sigma^2\mathbf{I}) \\
\boldsymbol{\mu}_t &\sim \begin{cases} \boldsymbol{\mu}_{0,h_t} 1_{\{\boldsymbol{\mu}_t \in \tilde{\mathcal{S}}\}} & z_t = 1 \\ \mathcal{TN}(\boldsymbol{\mu}_{0,h_t}, \sigma_\mu^2 \mathbf{I})_{\mathcal{S}} 1_{\{\boldsymbol{\mu}_t \in \mathcal{S}\}}, & z_t = 0 \end{cases} \\
z_t &\sim \text{Bern}(p) \\
h_t &\sim \text{Cat}\left(\frac{\pi_h [\mathbf{s}_t | \boldsymbol{\mu}_t, \sigma^2]^{z_t} [\boldsymbol{\mu}_t | \boldsymbol{\mu}_{0,h_t}, \sigma_\mu^2, \mathcal{S}]^{1-z_t}}{\sum_{\tilde{h}=1}^{H} \pi_{\tilde{h}} [\mathbf{s}_t | \boldsymbol{\mu}_t, \sigma^2]^{z_t} [\boldsymbol{\mu}_t | \boldsymbol{\mu}_{0,\tilde{h}}, \sigma_\mu^2, \mathcal{S}]^{1-z_t}} \right) \\
\pi_h &= v_h \prod_{\tilde{h} < h} (1 - v_{\tilde{h}}) \\
v_h &\sim \text{Beta}(1, \theta) \\
\boldsymbol{\mu}_{0,h} &\sim \text{Unif}(\tilde{\mathcal{S}}) \\
\theta &\sim \text{Gamma}(r, q) \\
p &\sim \text{Beta}(\alpha, \beta) \\
\sigma &\sim \text{Unif}(l_\sigma, u_\sigma) \\
\sigma_\mu &\sim \text{Unif}(l_{\sigma_\mu}, u_{\sigma_\mu})
\end{aligned}
$$

The concentration parameter θ affects the clustering in the Dirichlet process mixture: smaller values yield fewer clusters with more observations per cluster, whereas larger values yield more clusters with fewer observations per cluster. Note that the lines in this model statement pertaining to h_t, π_h, v_h, and $\boldsymbol{\mu}_{0,h}$ comprise the stick-breaking representation of the Dirichlet process mixture model, i.e.,

$$
\begin{aligned}
\boldsymbol{\mu}_{0,h} &\sim \mathbf{G} \\
\mathbf{G} &\sim \text{DP}(\theta, \mathbf{G}_0) \\
\mathbf{G}_0 &\sim \text{Unif}(\tilde{\mathcal{S}})
\end{aligned}
$$

Full conditional distributions

True locations ($\boldsymbol{\mu}_t$):

$$\begin{aligned}[\boldsymbol{\mu}_t|\cdot] &\propto [\mathbf{s}_t|\boldsymbol{\mu}_t,\sigma^2][\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t},z_t,\sigma^2_\mu,\mathcal{S},\tilde{\mathcal{S}}] \\ &\propto [\mathbf{s}_t|\boldsymbol{\mu}_t,\sigma^2][\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t},\tilde{\mathcal{S}}]^{z_t}[\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2_\mu,\mathcal{S}]^{1-z_t}.\end{aligned}$$

For $z_t = 0$,

$$\begin{aligned}[\boldsymbol{\mu}_t|\cdot] &\propto [\mathbf{s}_t|\boldsymbol{\mu}_t,\sigma^2][\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2_\mu,\mathcal{S}]^{1-z_t} \\ &\propto \exp\left\{-\frac{1}{2}\left(\mathbf{s}_t-\boldsymbol{\mu}_t\right)'\left(\sigma^2\mathbf{I}\right)^{-1}\left(\mathbf{s}_t-\boldsymbol{\mu}_t\right)\right\}\exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_t-\boldsymbol{\mu}_{0,h_t}\right)'\left(\sigma^2_\mu\mathbf{I}\right)^{-1}\left(\boldsymbol{\mu}_t-\boldsymbol{\mu}_{0,h_t}\right)\right\}1_{\{\boldsymbol{\mu}_t\in\mathcal{S}\}} \\ &\propto \exp\left\{-\frac{1}{2}\left(\mathbf{s}_t'\left(\sigma^2\mathbf{I}\right)^{-1}\mathbf{s}_t-2\mathbf{s}_t'\left(\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{\mu}_t+\boldsymbol{\mu}_t'\left(\sigma^2\mathbf{I}\right)^{-1}\boldsymbol{\mu}_t\right)\right\}\times \\ &\qquad \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_t'\left(\sigma^2_\mu\mathbf{I}\right)^{-1}\boldsymbol{\mu}_t-2\boldsymbol{\mu}_t'\left(\sigma^2_\mu\mathbf{I}\right)^{-1}\boldsymbol{\mu}_{0,h_t}+\boldsymbol{\mu}_{0,h_t}'\left(\sigma^2_\mu\mathbf{I}\right)^{-1}\boldsymbol{\mu}_{0,h_t}\right)\right\}1_{\{\boldsymbol{\mu}_t\in\mathcal{S}\}} \\ &\propto \exp\left\{-\frac{1}{2}\left(-2\left(\mathbf{s}_t'\left(\sigma^2\mathbf{I}\right)^{-1}+\boldsymbol{\mu}_{0,h_t}'\left(\sigma^2_\mu\mathbf{I}\right)^{-1}\right)\boldsymbol{\mu}_t+\boldsymbol{\mu}_t'\left(\left(\sigma^2\mathbf{I}\right)^{-1}+\left(\sigma^2_\mu\mathbf{I}\right)^{-1}\right)\boldsymbol{\mu}_t\right)\right\}1_{\{\boldsymbol{\mu}_t\in\mathcal{S}\}} \\ &= \mathcal{N}(\mathbf{A}^{-1}\mathbf{b},\mathbf{A}^{-1})1_{\{\boldsymbol{\mu}_t\in\mathcal{S}\}}\end{aligned}$$

where $\mathbf{A} = \left(\sigma^2\mathbf{I}\right)^{-1}+\left(\sigma^2_\mu\mathbf{I}\right)^{-1}$ and $\mathbf{b} = \mathbf{s}_t'\left(\sigma^2\mathbf{I}\right)^{-1}+\boldsymbol{\mu}_{0,h_t}'\left(\sigma^2_\mu\mathbf{I}\right)^{-1}$. Note that proposed values for $\boldsymbol{\mu}_t$ not in $\mathcal{S}$ are rejected, i.e., $[\boldsymbol{\mu}_t|\cdot] = \mathcal{TN}(\mathbf{A}^{-1}\mathbf{b},\mathbf{A}^{-1})_{\mathcal{S}}$.

For $z_t = 1$, recall that the true location is the location of the haul-out site (i.e., $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{0,h_t}$). Consequently, the update for $\boldsymbol{\mu}_t$ when $z_t = 1$ occurs during the updates for $\boldsymbol{\mu}_{0,h}$ and h_t.

Haul-out site locations ($\boldsymbol{\mu}_{0,h}$):

$$\begin{aligned}[\boldsymbol{\mu}_{0,h}|\cdot] &\propto \prod_{\{t:h_t=h\}}[\mathbf{s}_t|\boldsymbol{\mu}_t,\sigma^2]^{z_t}[\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h},\sigma^2_\mu,\mathcal{S}]^{1-z_t}[\boldsymbol{\mu}_{0,h}] \\ &\propto \prod_{\{t:h_t=h\}}\mathcal{N}(\mathbf{s}_t|\boldsymbol{\mu}_{0,h},\sigma^2)^{z_t}\mathcal{TN}(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h},\sigma^2_\mu)^{1-z_t}_{\mathcal{S}}1_{\{\boldsymbol{\mu}_{0,h}\in\tilde{\mathcal{S}}\}}\end{aligned}$$

The update for $\boldsymbol{\mu}_{0,h}$ proceeds using Metropolis-Hastings and proposals for $\boldsymbol{\mu}_{0,h}$ not in $\tilde{\mathcal{S}}$ are automatically rejected. Note that the product is over all $\boldsymbol{\mu}_t$ that are members of $\boldsymbol{\mu}_{0,h}$, and that $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{0,h_t}$ when $z_t = 1$. For each of the unoccupied haul-out locations (i.e., clusters with no members), sample $\boldsymbol{\mu}_{0,h}$ directly from the prior.

Haul-out indicator variable (z_t):

$$\begin{aligned}[z_t|\cdot] &\propto [\mathbf{s}_t|\boldsymbol{\mu}_t,\sigma^2]^{z_t}[\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2_\mu,\mathcal{S}]^{1-z_t}[z_t|p] \\ &\propto [\mathbf{s}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2]^{z_t}[\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2_\mu,\mathcal{S}]^{1-z_t}[z_t|p]\end{aligned}$$

For all $\boldsymbol{\mu}_t \notin \tilde{\mathcal{S}}$, let $z_t = 0$. For all $\boldsymbol{\mu}_t \in \tilde{\mathcal{S}}$,

$$[z_t|\cdot] = \text{Bern}(\tilde{p}),$$

where

$$\begin{aligned}\tilde{p} &= \frac{p[\mathbf{s}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2]}{p[\mathbf{s}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2]+(1-p)[\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2_\mu,\mathcal{S}]} \\ &= \frac{p\mathcal{N}(\mathbf{s}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2)}{p\mathcal{N}(\mathbf{s}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2)+(1-p)\mathcal{TN}(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2_\mu)_{\mathcal{S}}}.\end{aligned}$$

Haul-out classification variable (h_t):

$$
\begin{aligned}
[h_t|\cdot] &\sim \frac{\pi_h[\mathbf{s}_t|\boldsymbol{\mu}_t,\sigma^2]^{z_t}[\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2_\mu,\mathcal{S}]^{1-z_t}}{\sum_{\tilde{h}=1}^H \pi_{\tilde{h}}[\mathbf{s}_t|\boldsymbol{\mu}_t,\sigma^2]^{z_t}[\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,\tilde{h}},\sigma^2_\mu,\mathcal{S}]^{1-z_t}} \\
&\sim \frac{\pi_h\mathcal{N}(\mathbf{s}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2)^{z_t}\mathcal{TN}(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2_\mu)_{\mathcal{S}}^{1-z_t}}{\sum_{\tilde{h}=1}^H \pi_{\tilde{h}}\mathcal{N}(\mathbf{s}_t|\boldsymbol{\mu}_{0,\tilde{h}},\sigma^2)^{z_t}\mathcal{TN}(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,\tilde{h}},\sigma^2_\mu)_{\mathcal{S}}^{1-z_t}}.
\end{aligned}
$$

This update proceeds just as in multinomial sampling; see page 552 in Gelman et al. (2014).

Probability mass for haul-out location $\boldsymbol{\mu}_{0,h}$ (π_h):

$$
\pi_h = v_h \prod_{\tilde{h}<h}(1-v_{\tilde{h}}),
$$

where

$$
[v_h|\cdot] \sim \text{Beta}\left(1+n_h, \theta + \sum_{\tilde{h}=h+1}^H n_{\tilde{h}}\right), \text{ for } h=1,\ldots,H-1,
$$

and $v_H = 1$. This represents the stick-breaking construction of the Dirichlet process. The parameter n_h denotes the number of observations allocated to cluster h. Note that v_h is sampled in order of decreasing n_h, i.e., n_h is sorted largest to smallest and v_h is sampled in sequence. The probabilities π_h are calculated in order of decreasing n_h as well. See page 553 in Gelman et al. (2014) and Section 5.2 in Ishwaran and James (2001).

Dirichlet process concentration parameter (θ):

$$
[\theta|\cdot] \propto \text{Gamma}(r+H-1, q - \sum_{h=1}^{H-1}\log(1-v_h)).
$$

See page 553 in Gelman et al. (2014).

Probability of being hauled-out (p):

$$
\begin{aligned}
[p|\cdot] &\propto \prod_{t=1}^T [z_t|p][p] \\
&\propto \prod_{t=1}^T p^{z_t}(1-p)^{1-z_t}p^{\alpha-1}(1-p)^{\beta-1} \\
&\propto p^{\sum_{t=1}^T z_t}(1-p)^{\sum_{t=1}^T(1-z_t)}p^{\alpha-1}(1-p)^{\beta-1} \\
&= \text{Beta}\left(\sum_{t=1}^T z_t+\alpha, \sum_{t=1}^T(1-z_t)+\beta\right)
\end{aligned}
$$

Error in the observation process (σ):

$$
\begin{aligned}
[\sigma|\cdot] &\propto \prod_{t=1}^T [\mathbf{s}_t|\boldsymbol{\mu}_t,\sigma^2][\sigma] \\
&\propto \prod_{t=1}^T \mathcal{N}(\mathbf{s}_t|\boldsymbol{\mu}_t,\sigma^2)
\end{aligned}
$$

The update for σ proceeds using Metropolis-Hastings. Recall that $\boldsymbol{\mu}_t = \boldsymbol{\mu}_{0,h_t}$ when $z_t = 1$.

Homerange dispersion parameter (σ_μ):

$$\begin{aligned}
[\sigma_\mu|\cdot] &\propto \prod_{\{t:z_t=0\}} [\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t}, \sigma_\mu^2, \mathcal{S}][\sigma_\mu] \\
&\propto \prod_{\{t:z_t=0\}} \mathcal{TN}(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h_t}, \sigma_\mu^2)_{\mathcal{S}}
\end{aligned}$$

The update for σ_μ proceeds using Metropolis-Hastings. Note that the product is over all t for which $z_t = 0$.

References

Gelman, A., J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. 2014. Bayesian data analysis. CRC Press.

Ishwaran, H., and L.F. James. 2001. Gibbs sampling methods for stick-breaking priors. Journal of the American Statistical Association 96: 161–173.