

# HAUL-OUT SITE ESTIMATION USING A TRUNCATION APPROXIMATION OF A DIRICHLET PROCESS MIXTURE MODEL

Brian M. Brost

23 July 2015

## Implementation

The file `haulout.dp.mixture.sim.R` simulates data according to the model statement presented below, and `haulout.dp.mixture.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  $t = 1, \dots, T$ , be observed and true locations, respectively. Also let  $\boldsymbol{\mu}_{0,h} = (\mu_{0,1,h}, \mu_{0,2,h})'$ , for  $h = 1, \dots, H$ , 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  $\tilde{\mathcal{S}}$  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  $\tilde{\mathcal{S}}$  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).

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

The concentration parameter  $\theta$  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  $h_t$ ,  $\pi_h$ ,  $v_h$ , 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

*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}_{0,h},\sigma^2]^{z_t} [\mathbf{s}_t|\boldsymbol{\mu}_{0,h},\sigma^2,\sigma_\mu^2]^{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{N}(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{0,h},\sigma^2\mathbf{I}+\sigma_\mu^2\mathbf{I})^{1-z_t} 1_{\{\boldsymbol{\mu}_{0,h}\in\tilde{\mathcal{S}}\}} \\
&\propto \prod_{\{t:h_t=h\}} \exp\left\{-\frac{1}{2}(\mathbf{s}_t-\boldsymbol{\mu}_{0,h})'(\sigma^2\mathbf{I})^{-1}(\mathbf{s}_t-\boldsymbol{\mu}_{0,h})\right\}^{z_t} \\
&\quad \exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_t-\boldsymbol{\mu}_{0,h})'(\sigma^2\mathbf{I}+\sigma_\mu^2\mathbf{I})^{-1}(\boldsymbol{\mu}_t-\boldsymbol{\mu}_{0,h})\right\}^{1-z_t} 1_{\{\boldsymbol{\mu}_{0,h}\in\tilde{\mathcal{S}}\}} \\
&\propto \prod_{\{t:h_t=h\}} \exp\left\{-\frac{1}{2}\left(\mathbf{s}_t'(\sigma^2\mathbf{I})^{-1}\mathbf{s}_t-2\mathbf{s}_t'(\sigma^2\mathbf{I})^{-1}\boldsymbol{\mu}_{0,h}+\boldsymbol{\mu}_{0,h}'(\sigma^2\mathbf{I})^{-1}\boldsymbol{\mu}_{0,h}\right)\right\}^{z_t} \times \\
&\quad \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_t'(\sigma^2\mathbf{I}+\sigma_\mu^2\mathbf{I})^{-1}\boldsymbol{\mu}_t-2\boldsymbol{\mu}_t'(\sigma^2\mathbf{I}+\sigma_\mu^2\mathbf{I})^{-1}\boldsymbol{\mu}_{0,h}+\boldsymbol{\mu}_{0,h}'(\sigma^2\mathbf{I}+\sigma_\mu^2\mathbf{I})^{-1}\boldsymbol{\mu}_{0,h}\right)\right\}^{1-z_t} \times \\
&\quad 1_{\{\boldsymbol{\mu}_{0,h}\in\tilde{\mathcal{S}}\}} \\
&\propto \prod_{\{t:h_t=h\}} \exp\left\{-\frac{1}{2}\left(-2\left(\mathbf{s}_t'(\sigma^2\mathbf{I})^{-1}1_{\{z_t=1\}}+\boldsymbol{\mu}_t'(\sigma^2\mathbf{I}+\sigma_\mu^2\mathbf{I})^{-1}1_{\{z_t=0\}}\right)\boldsymbol{\mu}_{0,h}+\right.\right. \\
&\quad \left.\left.\boldsymbol{\mu}_{0,h}'\left((\sigma^2\mathbf{I})^{-1}1_{\{z_t=1\}}+(\sigma^2\mathbf{I}+\sigma_\mu^2\mathbf{I})^{-1}1_{\{z_t=0\}}\right)\boldsymbol{\mu}_{0,h}\right)\right\} 1_{\{\boldsymbol{\mu}_{0,h}\in\tilde{\mathcal{S}}\}} \\
&\propto \exp\left\{-\frac{1}{2}\left(-2\left(\sum_{\{t:h_t=0,z_t=1\}}\mathbf{s}_t'(\sigma^2\mathbf{I})^{-1}+\sum_{\{t:h_t=0,z_t=0\}}\boldsymbol{\mu}_t'(\sigma^2\mathbf{I}+\sigma_\mu^2\mathbf{I})^{-1}\right)\boldsymbol{\mu}_{0,h}+\right.\right. \\
&\quad \left.\left.\boldsymbol{\mu}_{0,h}'\left(n_{h,1}(\sigma^2\mathbf{I})^{-1}+n_{h,0}(\sigma^2\mathbf{I}+\sigma_\mu^2\mathbf{I})^{-1}\right)\boldsymbol{\mu}_{0,h}\right)\right\} 1_{\{\boldsymbol{\mu}_{0,h}\in\tilde{\mathcal{S}}\}} \\
&= \mathcal{N}(\mathbf{A}^{-1}\mathbf{b},\mathbf{A}^{-1})1_{\{\boldsymbol{\mu}_t\in\mathcal{S}\}}
\end{aligned}$$

where  $\mathbf{A} = n_{h,1}(\sigma^2\mathbf{I})^{-1} + n_{h,0}(\sigma^2\mathbf{I} + \sigma_\mu^2\mathbf{I})^{-1}$  and  $\mathbf{b} = \sum_{\{t:h_t=0,z_t=1\}} \mathbf{s}_t'(\sigma^2\mathbf{I})^{-1} + \sum_{\{t:h_t=0,z_t=0\}} \boldsymbol{\mu}_t'(\sigma^2\mathbf{I} + \sigma_\mu^2\mathbf{I})^{-1}$ . Note that the product is over all  $\mathbf{s}_t$  that are members of  $\boldsymbol{\mu}_{0,h}$ ,  $n_{h,1}$  is the number of members in cluster  $h$  where  $z_t = 1$ , and  $n_{h,0}$  is the number of members in cluster  $h$  where  $z_t = 0$ . Proposed values for  $\boldsymbol{\mu}_{0,h}$  not in  $\tilde{\mathcal{S}}$  are rejected, i.e.,  $[\boldsymbol{\mu}_{0,h}|\cdot] = \mathcal{TN}(\mathbf{A}^{-1}\mathbf{b},\mathbf{A}^{-1})_{\tilde{\mathcal{S}}}$ . For each of the unoccupied haul-out locations (i.e., clusters with no members),  $\boldsymbol{\mu}_{0,h}$  is sampled directly from the prior.

*Haul-out indicator variable* ( $z_t$ ):

$$\begin{aligned}
[z_t|\cdot] &\propto [\mathbf{s}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2]^{z_t} [\mathbf{s}_t|\boldsymbol{\mu}_{0,h_t},\sigma^2,\sigma_\mu^2]^{1-z_t} [z_t|p] \\
&= \text{Bern}(\tilde{p}),
\end{aligned}$$

where

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

*Haul-out classification variable* ( $h_t$ ):

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

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

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

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

where

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

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  $\pi_h$  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 ( $\theta$ ):*

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

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 ( $\sigma$ ):*

$$\begin{aligned}
[\sigma|\cdot] &\propto \prod_{t=1}^T [\mathbf{s}_t | \boldsymbol{\mu}_{0,h_t}, \sigma^2]^{z_t} [\mathbf{s}_t | \boldsymbol{\mu}_{0,h_t}, \sigma^2, \sigma_\mu^2]^{1-z_t} [\sigma] \\
&\propto \prod_{t=1}^T \mathcal{N}(\mathbf{s}_t | \boldsymbol{\mu}_{0,h_t}, \sigma^2 \mathbf{I})^{z_t} \mathcal{N}(\mathbf{s}_t | \boldsymbol{\mu}_{0,h_t}, \sigma^2 \mathbf{I} + \sigma_\mu^2 \mathbf{I})^{1-z_t}
\end{aligned}$$

The update for  $\sigma$  proceeds using Metropolis-Hastings.

*Homerange dispersion parameter* ( $\sigma_\mu$ ):

$$\begin{aligned} [\sigma_\mu | \cdot] &\propto \prod_{\{t: z_t=0\}} [\mathbf{s}_t | \boldsymbol{\mu}_{0, h_t}, \sigma^2, \sigma_\mu^2] [\sigma_\mu] \\ &\propto \prod_{\{t: z_t=0\}} \mathcal{N}(\mathbf{s}_t | \boldsymbol{\mu}_{0, h_t}, \sigma^2 \mathbf{I} + \sigma_\mu^2 \mathbf{I}) \end{aligned}$$

The update for  $\sigma_\mu$  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.