Bayesian model selection

for

Markov chains

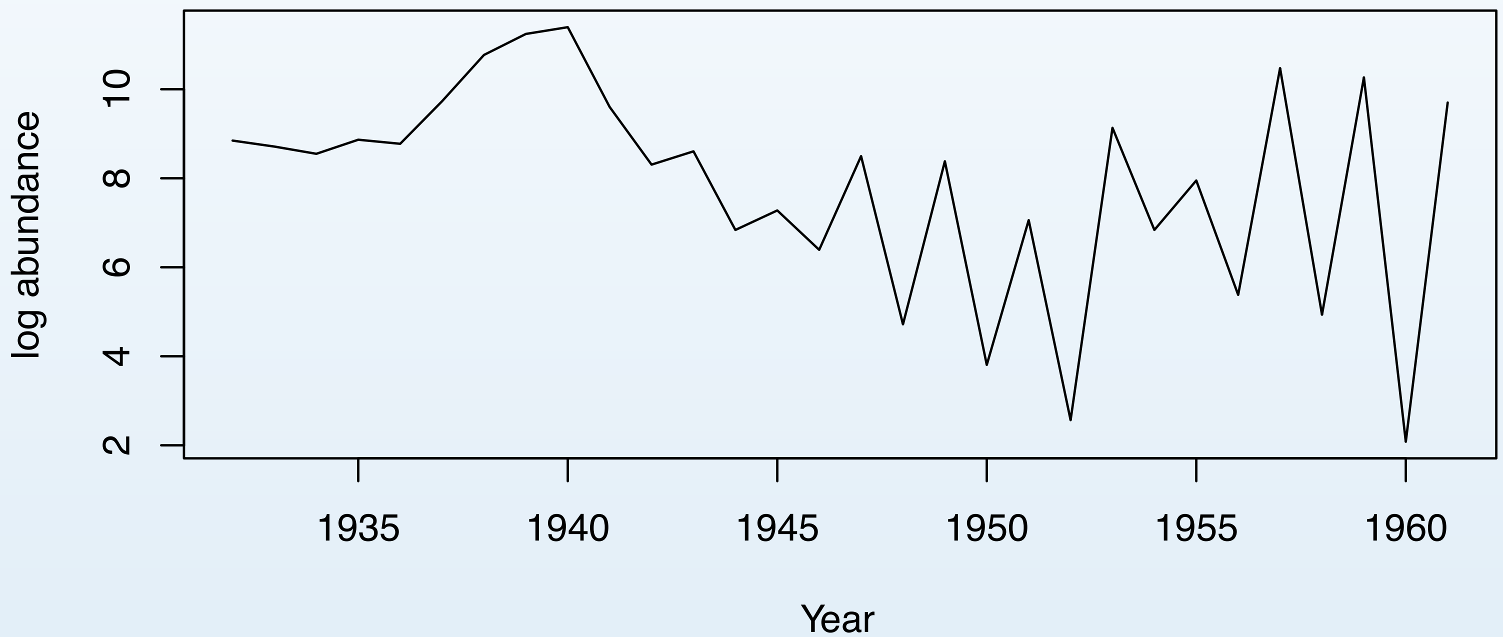
using

sparse probability vectors

Matthew Heiner[1], Athanasios Kottas[1], and Stephan Munch[2]

1. Department of Applied Mathematics and Statistics, University of California, Santa Cruz, California, USA
2. Fisheries Ecology Division, Southwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Santa Cruz, California, USA

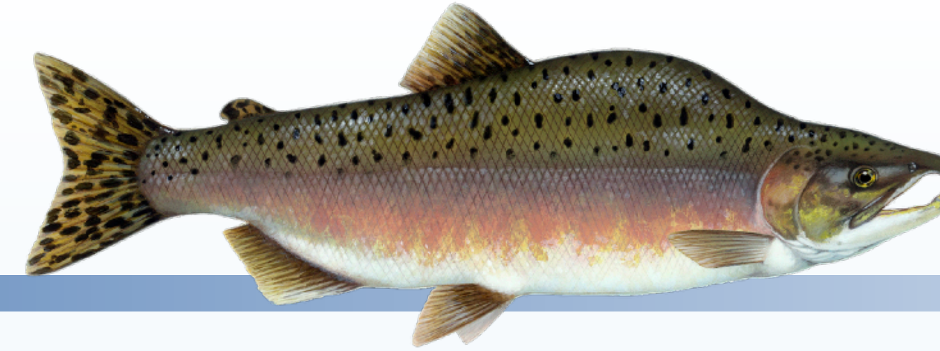
Pink Salmon time series

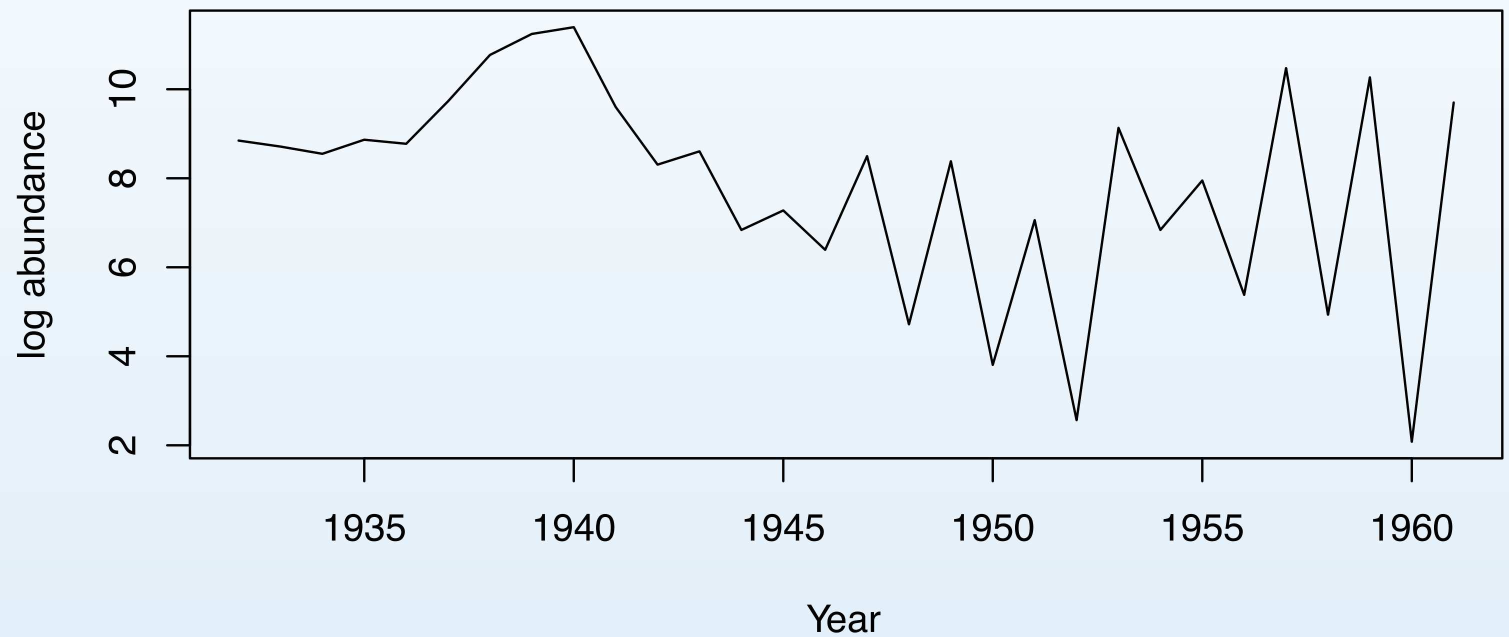


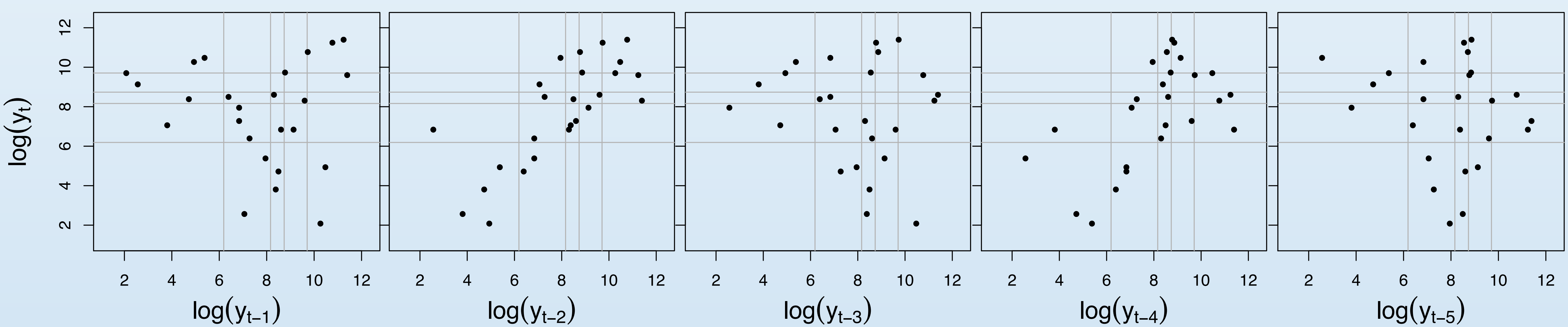
Data: Alaska Fisheries Science Center (2018), "AFSC/ABL: Pink salmon data collected at Sashin Creek Weir 1934-2002," NOAA National Centers for Environmental Information, https://inport.nmfs.noaa.gov/ inport/item/17256.

Image: U.S. Fish and Wildlife Service

Pink Salmon time series







Data: Alaska Fisheries Science Center (2018), "AFSC/ABL: Pink salmon data collected at Sashin Creek Weir 1934-2002," NOAA National Centers for Environmental Information, https://inport.nmfs.noaa.gov/ inport/item/17256.

Image: U.S. Fish and Wildlife Service

Mixture transition distribution Raftery (1985)

Mixture transition distribution

Raftery (1985)

Example with two states and three lags:

Mixture transition distribution

Raftery (1985)

Example with two states and three lags:

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} \\ q_{2,1} & q_{2,2} \end{pmatrix}$$

first order
transition matrix

Mixture transition distribution

Raftery (1985)

Example with two states and three lags:

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} \\ q_{2,1} & q_{2,2} \end{pmatrix}$$

first order
transition matrix

$$\Pr(s_t = 2 \mid s_{t-1} = 1, s_{t-2} = 1, s_{t-3} = 2)$$

Mixture transition distribution

Raftery (1985)

Example with two states and three lags:

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} \\ q_{2,1} & q_{2,2} \end{pmatrix}$$

first order
transition matrix

$$\Pr(s_t = 2 \mid s_{t-1} = 1, s_{t-2} = 1, s_{t-3} = 2)$$

$$= \lambda_1 q_{1,2}$$

Mixture transition distribution

Raftery (1985)

Example with two states and three lags:

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} \\ q_{2,1} & q_{2,2} \end{pmatrix}$$

first order
transition matrix

$$\Pr(s_t = 2 \mid s_{t-1} = 1, s_{t-2} = 1, s_{t-3} = 2)$$

$$= \lambda_1 q_{1,2} + \lambda_2 q_{1,2}$$

Mixture transition distribution

Raftery (1985)

Example with two states and three lags:

$$Q = \begin{pmatrix} q_{1,1} & q_{1,2} \\ q_{2,1} & q_{2,2} \end{pmatrix}$$

first order transition matrix

$$\Pr(s_t = 2 \mid s_{t-1} = 1, s_{t-2} = 1, s_{t-3} = 2)$$

$$= \lambda_1 q_{1,2} + \lambda_2 q_{1,2} + \lambda_3 q_{2,2}$$

Sparse probability vectors

Sparse Dirichlet mixture (SDM) prior

$$p(\boldsymbol{\theta}) \propto \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha}) \times \sum_{k=1}^{K} \theta_k^{\beta}, \quad \beta > 1$$

Sparse probability vectors

Sparse Dirichlet mixture (SDM) prior

$$p(\boldsymbol{\theta}) \propto \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha}) \times \sum_{k=1}^{K} \theta_k^{\beta}, \quad \beta > 1$$

$$p_{\mathrm{SDM}}(\boldsymbol{\theta}; \boldsymbol{\alpha}, \beta) = \sum_{k=1}^{K} \frac{w_k}{\sum_{j=1}^{K} w_j} \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha} + \beta \mathbf{e}_k)$$

$$w_k = \prod_{j=1}^{K} \Gamma(\alpha_j + \beta 1_{(j=k)})$$

Sparse probability vectors

Sparse Dirichlet mixture (SDM) prior

$$p(\boldsymbol{\theta}) \propto \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha}) \times \sum_{k=1}^{K} \theta_k^{\beta}\,, \quad \beta > 1$$

$$p_{\mathrm{SDM}}(\boldsymbol{\theta}; \boldsymbol{\alpha}, \beta) = \sum_{k=1}^{K} \frac{w_k}{\sum_{j=1}^{K} w_j} \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha} + \beta \mathbf{e}_k)$$

$$w_k = \prod_{j=1}^{K} \Gamma(\alpha_j + \beta 1_{(j=k)})$$

$$p(\boldsymbol{\theta} \mid \boldsymbol{n}) \quad \propto \quad \prod_{k=1}^{K} \theta_k^{n_k} \times \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha}) \times \sum_{k=1}^{K} \theta_k^{\beta} \quad \propto \quad \mathrm{Dir}(\boldsymbol{\theta}; \boldsymbol{\alpha} + \boldsymbol{n}) \times \sum_{k=1}^{K} \theta_k^{\beta}$$

Sparse probability vectors

Stick-breaking mixture (SBM) prior

(Connor and Mosimann, 1969)

Sparse probability vectors

Stick-breaking mixture (SBM) prior

$$\theta_1 = X_1,$$

(Connor and Mosimann, 1969)

Sparse probability vectors

Stick-breaking mixture (SBM) prior

$$\theta_1 = X_1, \;\; \theta_k = X_k \prod_{j=1}^{k-1} (1 - X_j) \text{ for } k = 2, \ldots, K-1,$$

(Connor and Mosimann, 1969)

Sparse probability vectors

Stick-breaking mixture (SBM) prior

$$\theta_1 = X_1,\ \ \theta_k = X_k \prod_{j=1}^{k-1}(1 - X_j) \text{ for } k = 2, \ldots, K-1,\ \ \theta_K = 1 \cdot \prod_{j=1}^{K-1}(1 - X_j)\,,$$

(Connor and Mosimann, 1969)

Sparse probability vectors

Stick-breaking mixture (SBM) prior

$$\theta_1 = X_1, \;\; \theta_k = X_k \prod_{j=1}^{k-1} (1 - X_j) \text{ for } k = 2, \ldots, K-1, \;\; \theta_K = 1 \cdot \prod_{j=1}^{K-1} (1 - X_j) \,,$$

(Connor and Mosimann, 1969)

$$X_k \overset{\text{ind.}}{\sim} \pi \text{Beta}(1, \eta) + (1 - \pi) \text{Beta}(\gamma, \delta)$$

Sparse probability vectors

Stick-breaking mixture (SBM) prior

$$\theta_1 = X_1, \;\; \theta_k = X_k \prod_{j=1}^{k-1} (1 - X_j) \text{ for } k = 2, \ldots, K-1, \;\; \theta_K = 1 \cdot \prod_{j=1}^{K-1} (1 - X_j)\,,$$

(Connor and Mosimann, 1969)

$$X_k \overset{\text{ind.}}{\sim} \pi \mathrm{Beta}(1, \eta) + (1 - \pi) \mathrm{Beta}(\gamma, \delta)$$

Smaller probabilities

Sparse probability vectors

Stick-breaking mixture (SBM) prior

$$\theta_1 = X_1,\ \ \theta_k = X_k \prod_{j=1}^{k-1}(1 - X_j) \text{ for } k = 2, \ldots, K-1,\ \ \theta_K = 1 \cdot \prod_{j=1}^{K-1}(1 - X_j)\,,$$

(Connor and Mosimann, 1969)

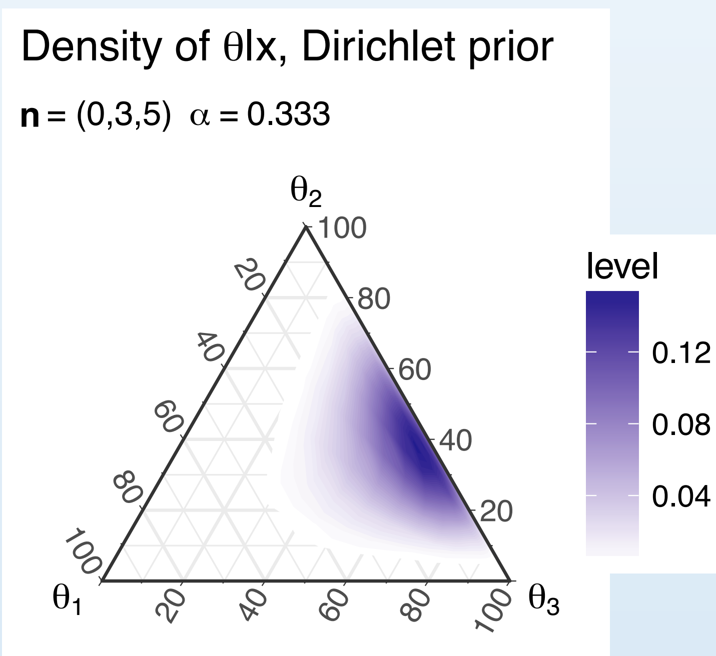
$$X_k \overset{\text{ind.}}{\sim} \pi \mathrm{Beta}(1, \eta) + (1 - \pi)\mathrm{Beta}(\gamma, \delta)$$

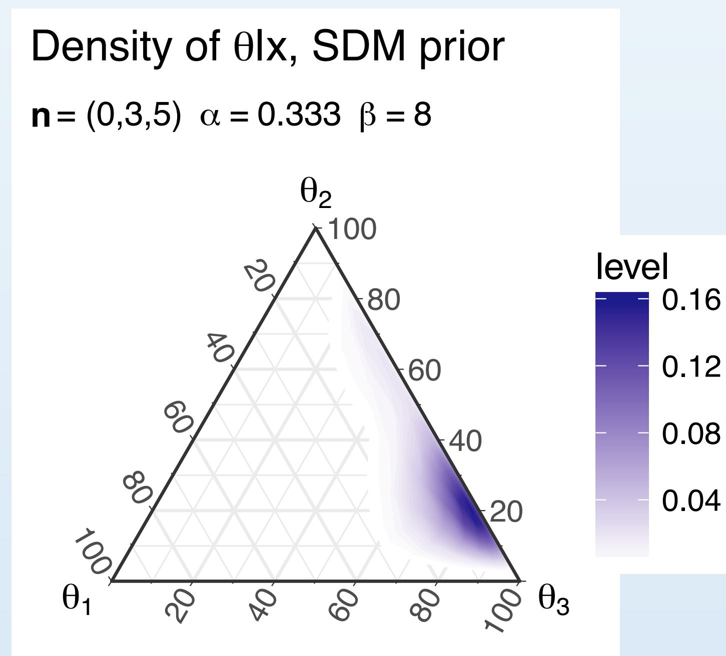
Smaller probabilities

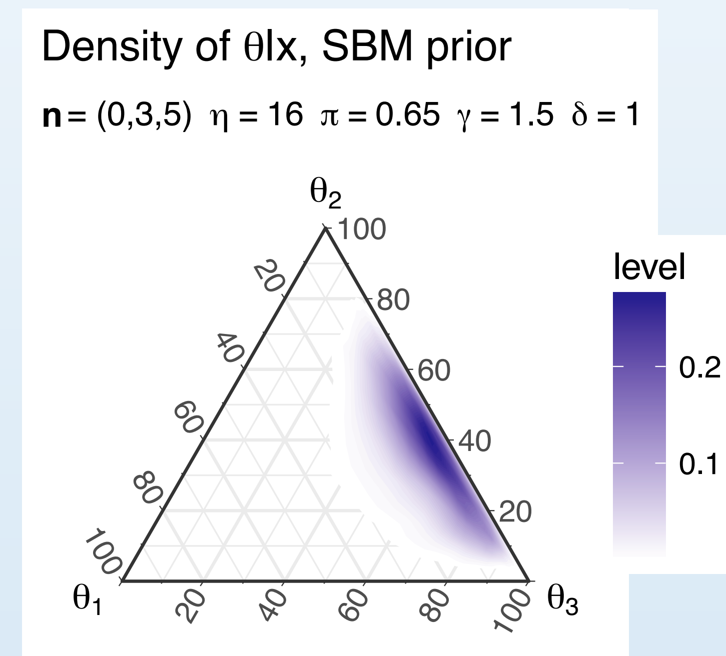
Larger probabilities

Sparse probability vectors

Multinomial data example







MTD extension

Let $J<R$ be a positive integer representing the highest order of mixing transition tensor with maximum lag R.

$$\Pr(s_t = i_0 \mid s_{t-1} = i_1, s_{t-2} = i_2, \dots, s_{t-R} = i_R, \boldsymbol{\Lambda}, \{\boldsymbol{\lambda}^{(j)}\}_{j=1}^{J}, \{\boldsymbol{Q}^{(j)}\}_{j=1}^{J}) = (\boldsymbol{\Omega})_{i_R, i_{R-1}, \dots, i_2, i_1, i_0}$$

MTD extension

Let $J<R$ be a positive integer representing the highest order of mixing transition tensor with maximum lag R.

$$\Pr(s_t = i_0 \mid s_{t-1} = i_1, s_{t-2} = i_2, \ldots, s_{t-R} = i_R, \boldsymbol{\Lambda}, \{\boldsymbol{\lambda}^{(j)}\}_{j=1}^{J}, \{\boldsymbol{Q}^{(j)}\}_{j=1}^{J}) = (\boldsymbol{\Omega})_{i_R, i_{R-1}, \ldots, i_2, i_1, i_0}$$

$$\equiv \Lambda_1 \sum_{\ell=1}^{R} \lambda_\ell^{(1)} Q^{(1)}(s_t = i_0 \mid s_{t-\ell} = i_\ell) +$$

MTD extension

Let *J*<*R* be a positive integer representing the highest order of mixing transition tensor with maximum lag *R*.

$$\Pr(s_t = i_0 \mid s_{t-1} = i_1, s_{t-2} = i_2, \ldots, s_{t-R} = i_R, \boldsymbol{\Lambda}, \{\boldsymbol{\lambda}^{(j)}\}_{j=1}^J, \{\boldsymbol{Q}^{(j)}\}_{j=1}^J) = (\boldsymbol{\Omega})_{i_R, i_{R-1}, \ldots, i_2, i_1, i_0}$$

$$\equiv \Lambda_1 \sum_{\ell=1}^{R} \lambda_\ell^{(1)} Q^{(1)}(s_t = i_0 \mid s_{t-\ell} = i_\ell) +$$

$$+\Lambda_2 \sum\sum_{\ell_1 < \ell_2} \lambda_{(\ell_1,\ell_2)}^{(2)} Q^{(2)}(s_t = i_0 \mid s_{t-\ell_1} = i_{\ell_1}, s_{t-\ell_2} = i_{\ell_2}) + \ldots +$$

MTD extension

Let $J<R$ be a positive integer representing the highest order of mixing transition tensor with maximum lag R.

$$\Pr(s_t = i_0 \mid s_{t-1} = i_1, s_{t-2} = i_2, \ldots, s_{t-R} = i_R, \boldsymbol{\Lambda}, \{\boldsymbol{\lambda}^{(j)}\}_{j=1}^{J}, \{\boldsymbol{Q}^{(j)}\}_{j=1}^{J}) = (\boldsymbol{\Omega})_{i_R, i_{R-1}, \ldots, i_2, i_1, i_0}$$

$$\equiv \Lambda_1 \sum_{\ell=1}^{R} \lambda_{\ell}^{(1)} Q^{(1)}(s_t = i_0 \mid s_{t-\ell} = i_\ell) +$$

$$+\Lambda_2 \sum_{\ell_1<\ell_2}\sum \lambda_{(\ell_1,\ell_2)}^{(2)} Q^{(2)}(s_t = i_0 \mid s_{t-\ell_1} = i_{\ell_1}, s_{t-\ell_2} = i_{\ell_2}) + \ldots +$$

$$+\Lambda_J \sum_{\ell_1<\ldots<\ell_J} \ldots \sum \lambda_{(\ell_1,\ldots,\ell_J)}^{(J)} Q^{(J)}(s_t = i_0 \mid s_{t-\ell_1} = i_{\ell_1}, \ldots, s_{t-\ell_J} = i_{\ell_J})$$

MTD extension

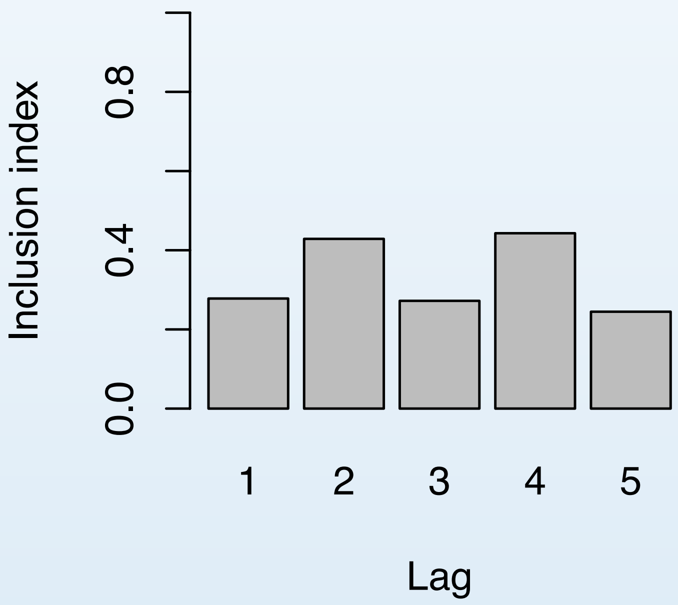
Free parameter count:

K	R	J	Λ	λ	Q	total	unrestricted
5	3	2	1	4	120	125	500
5	4	3	2	11	620	633	2,500
2	7	4	3	94	30	127	128
3	5	3	2	22	78	102	486
7	5	3	2	22	2,394	2,418	100,842
7	5	2	1	13	336	350	100,842

Pink Salmon time series

Lag inclusion inference

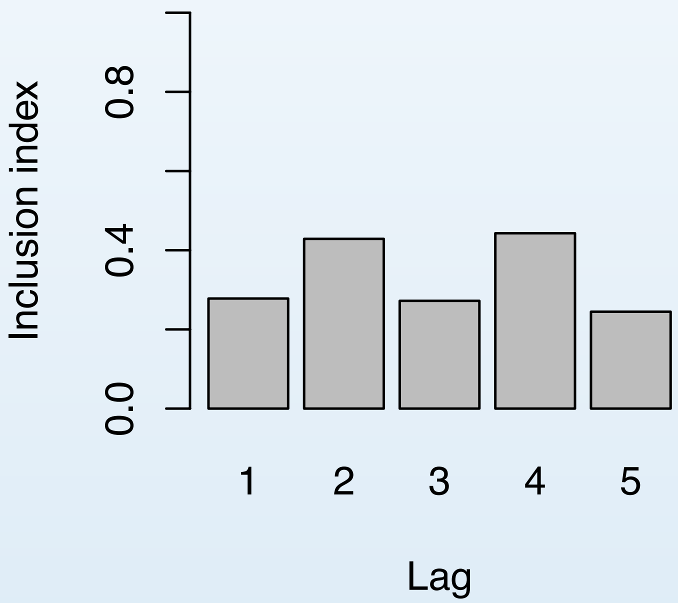
Dirichlet on mixing parameters



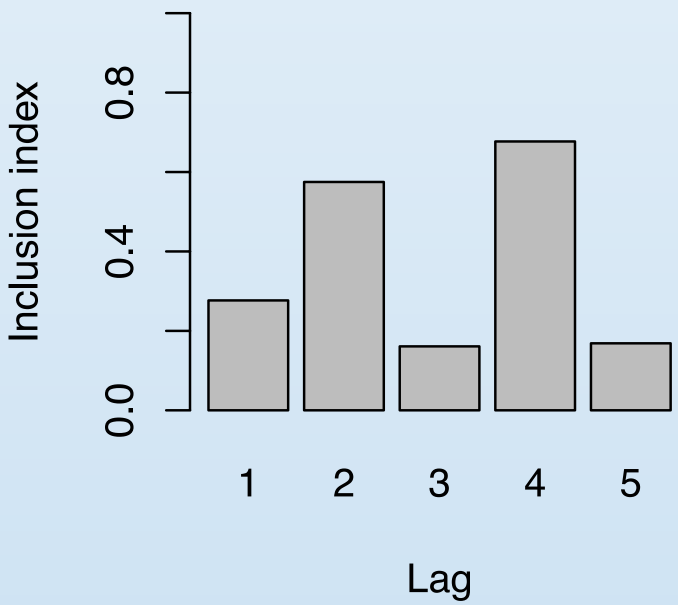
Pink Salmon time series

Lag inclusion inference

Dirichlet on mixing parameters



SDM on mixing parameters

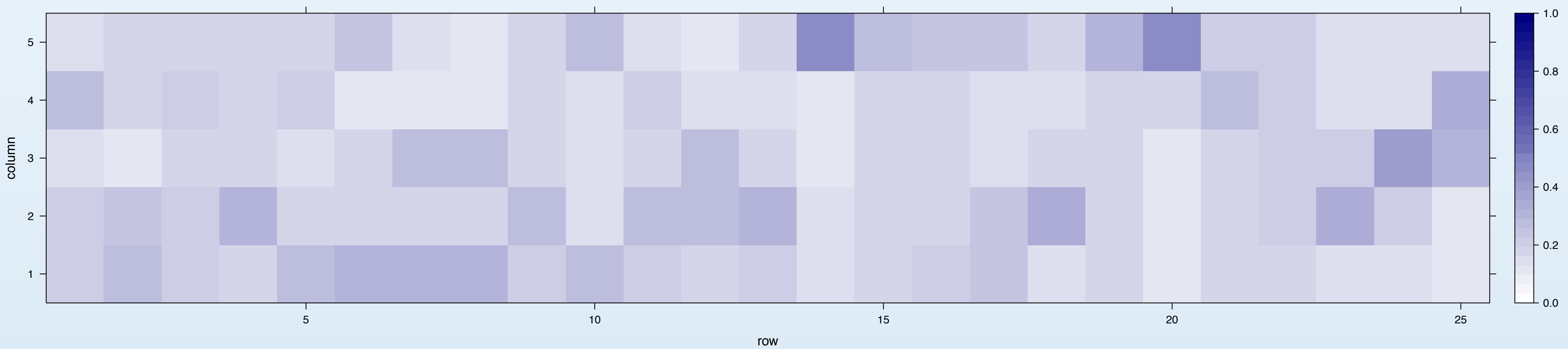


Pink Salmon time series

Posterior mean of $\boldsymbol{Q}$:

Second order with lags 2 and 4 most used

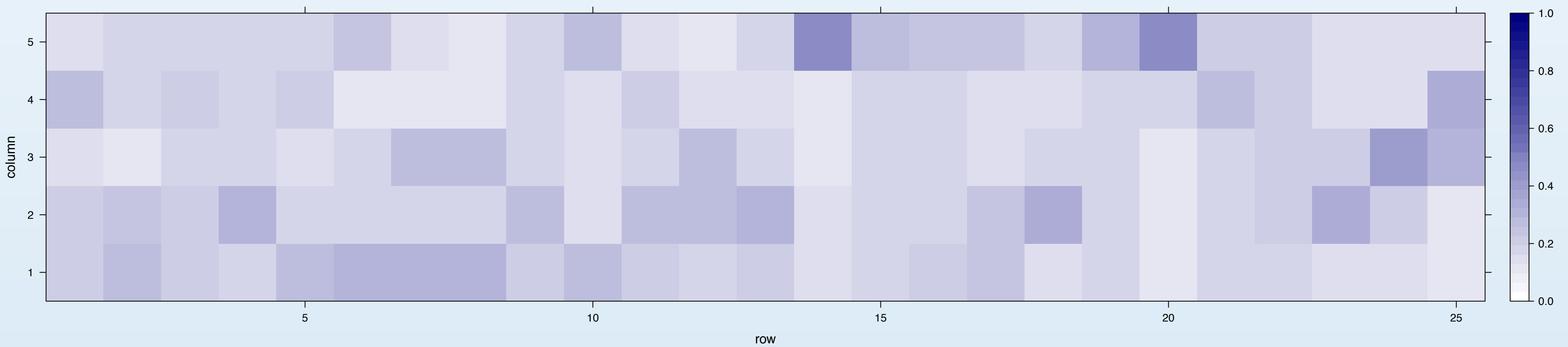
Dirichlet on mixing parameters (.67 for order 2)



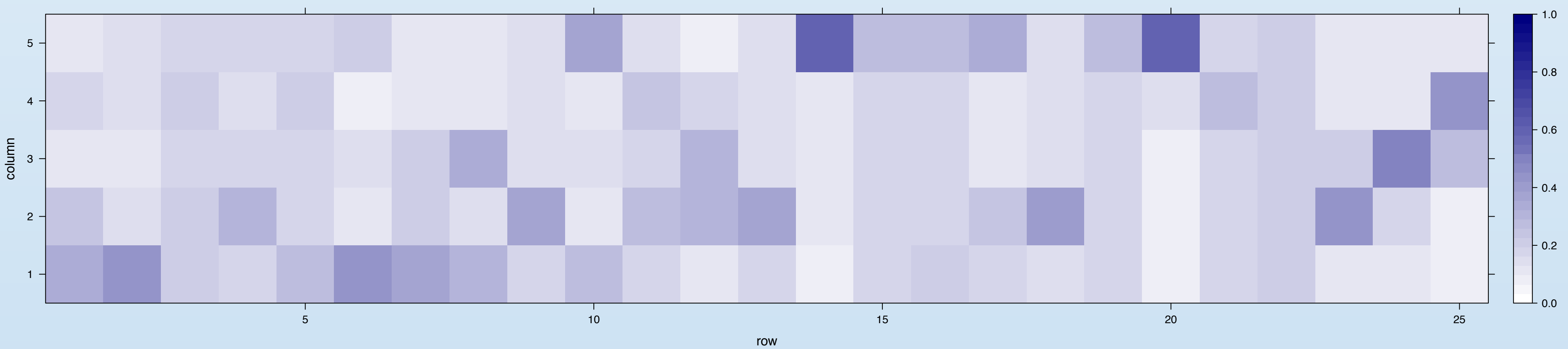
Pink Salmon time series

Posterior mean of $\boldsymbol{Q}$:
Second order with lags 2 and 4 most used

Dirichlet on mixing parameters (.67 for order 2)



SDM on mixing parameters (.86 for order 2)



Acknowledgements:

San Francisco Bay Area Chapter of ASA Student Travel Award for JSM 2018.

The work of the first and the second author was supported in part by the National Science Foundation under award DMS 1310438.

References:

Berchtold, André, and Adrian E. Raftery. "The mixture transition distribution model for high-order Markov chains and non-Gaussian time series.” *Statistical Science* (2002): 328-356.

Connor, Robert J., and James E. Mosimann. "Concepts of independence for proportions with a generalization of the Dirichlet distribution." *Journal of the American Statistical Association* 64.325 (1969): 194-206.

Raftery, Adrian E. "A model for high-order Markov chains." *Journal of the Royal Statistical Society. Series B (Methodological)* (1985): 528-539.

MTD extension

Hierarchical representation

$$(\boldsymbol{Q}^{(j)})_{i_j,i_{j-1},\ldots,i_1,\cdot} \sim \mathrm{Dir}(\boldsymbol{\alpha}_{\boldsymbol{Q}^{(j)}}) \quad \forall \quad (i_j, i_{j-1}, \ldots, i_1) \in \{1, \ldots, K\}^j \,,$$

$$\boldsymbol{\Lambda} \sim \mathrm{Dir}(\boldsymbol{\alpha}_\Lambda)\,, \qquad \boldsymbol{\lambda}^{(j)} \sim \mathrm{Dir}(\boldsymbol{\alpha}_{\lambda^{(j)}})\,,$$

$$\Pr(Z_t = j \mid \boldsymbol{\Lambda}) = \Lambda_j \,, \qquad \Pr(\boldsymbol{z}_t = (\ell_1, \ldots, \ell_j) \mid Z_t = j, \boldsymbol{\lambda}^{(j)}) = \lambda^{(j)}_{(\ell_1,\ldots,\ell_j)} \,,$$

$$\Pr(s_t = i_0 \,| s_{t-1} = i_1, s_{t-2} = i_2, \ldots, s_{t-R} = i_R, Z_t = j, z_t^{(j)} = (\ell_1, \ldots, \ell_j), \boldsymbol{Q}^{(j)}) = (\boldsymbol{Q}^{(j)})_{i_{\ell_j}, i_{\ell_{j-1}}, \ldots, i_{\ell_1}, i_0}$$