Slice Sampling Mixture Models

Maria Kalli, Jim E. Griffin, Stephen G. Walker
Statistics and Computing, 2011

presented by
Paul Kernfeld + Libin Sun

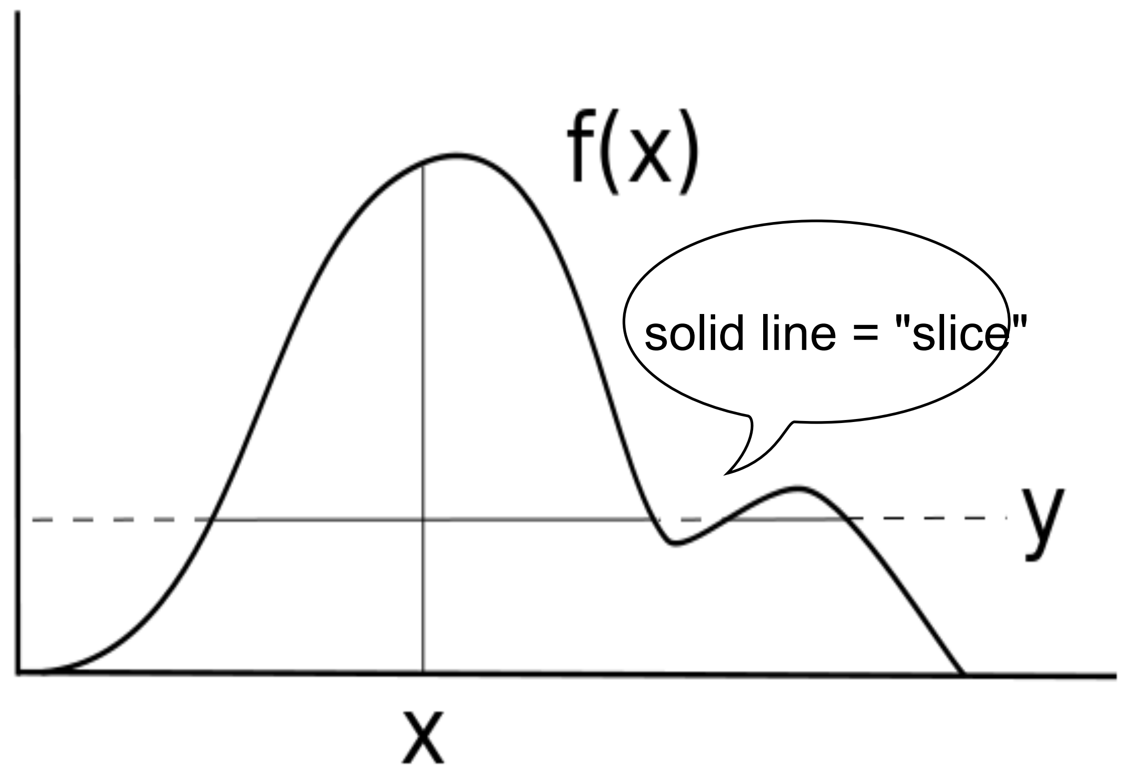
Overview

- slice sampling (Neal 2003)
- slice sampling for MDP (Walker 2007)
- slice-efficient sampler for MDP
- mixtures based on normalized weights
    - definition, properties
    - samplers
- comparison of different samplers

Background: Slice Sampling

univariate case: $f_i(x_i) \propto p(x_i | x_{-i})$

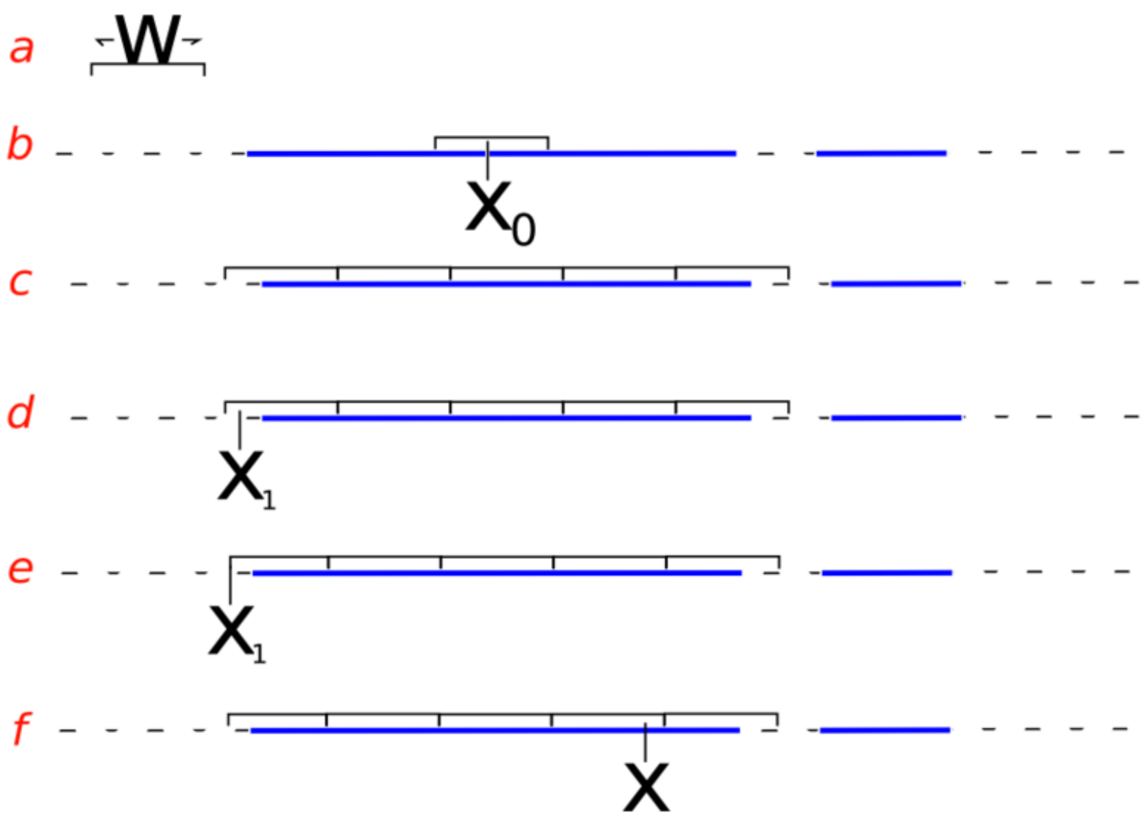
1. draw y~Unif(0,f(x))
2. draw x* from "slice"



Background: Slice Sampling

choosing the interval (Neal 2003):

1. stepping out
2. doubling



Slice Sampling for MDP

big idea: introduce latent variable *u*, consider the joint *f(y,u)* and slicing over *u*

$$P = \sum_{j=1}^{\infty} w_j \delta_{\phi_j}, \qquad w_1 = v_1, \qquad w_j = v_j \prod_{l<j} (1 - v_l)$$

$$f_{v,\mu,\sigma^2}(y) = \sum_{j=1}^{\infty} w_j \mathrm{N}(y; \mu_j, \sigma_j^2)$$

$$f_{v,\mu,\sigma^2}(y, u) = \sum_{j=1}^{\infty} \mathbf{1}(u < w_j) \mathrm{N}(y; \mu_j, \sigma_j^2).$$

Slice Sampling for MDP

$$f_{v,\mu,\sigma^2}(y,u) = \sum_{j=1}^{\infty} \mathbf{1}(u < w_j) \mathrm{N}(y; \mu_j, \sigma_j^2).$$

marginal of u: $f(u) = \sum_{j=1}^{\infty} \mathbf{1}(u < w_j)$ => staircases

marginal of y: $f_{v,\mu,\sigma^2}(y) = \sum_{j=1}^{\infty} w_j \mathrm{N}(y; \mu_j, \sigma_j^2)$

conditional of y|u: $f(y|u) = \dfrac{f(y,u)}{f(u)}$ ==> finite, equal weights

- is there a typo in the paper for N_u ?

Slice Sampling for MDP

latent variable *u* and *d*

- *u* to threshold *w*
- *d* is assignment variable (*z*)

$$f_{v,\mu,\sigma^2}(y,u,d) = \xi_d^{-1}\mathbf{1}(u < \xi_d) w_d \mathrm{N}(y; \mu_d, \sigma_d^2)$$

- set \xi_d = w_d: Walker 2007, Dunson 2008, etc
  - *u* and *w* correlated, slow mixing
  - simulate *w* when *u* changes

Slice-efficient Sampling for MDP

\xi: decreasing,
joint posterior:

$$\xi_j \propto \frac{E[w_j]}{1.5^j}$$

$$p(\mu_j, \sigma_j, v_j, d_i, u_i | y) \propto \prod_{i=1}^{n} \mathbf{1}(u_i < \xi_{d_i}) w_{d_i} / \xi_{d_i} \mathrm{N}(y_i ; \mu_{d_i}, \sigma_{d_i}^2).$$

1. $\pi(\mu_j, \sigma_j^2 | \cdots) \propto p_0(\mu_j, \sigma_j^2) \prod_{d_i = j} \mathrm{N}(y_i ; \mu_j, \sigma_j^2)$.
2. $\pi(v_j) \propto \mathrm{Be}(v_j ; a_j, b_j)$, where

$$a_j = 1 + \sum_{i=1}^{n} \mathbf{1}(d_i = j)$$

and

$$b_j = M + \sum_{i=1}^{n} \mathbf{1}(d_i > j).$$

3. $\pi(u_i | \cdots) \propto \mathbf{1}(0 < u_i < \xi_{d_i})$.
4. $\mathrm{P}(d_i = k | \cdots) \propto \mathbf{1}(k : \xi_k > u_i) w_k / \xi_k \mathrm{N}(y_i ; \mu_k, \sigma_k^2)$.

we can do this
because
{\xi} are fixed

Normalized Weights Mixtures

where: $$f(y) = \sum_{j=1}^{\infty} w_j K(y; \phi_j)$$

$$w_j = \lambda_j / \Lambda, \ \Lambda = \sum_{j=1}^{\infty} \lambda_j$$

$$\Lambda_m = \sum_{j=m+1}^{\infty} \lambda_j.$$

$\lambda_j \sim \pi_j(\lambda_j).$ ==> defines the properties

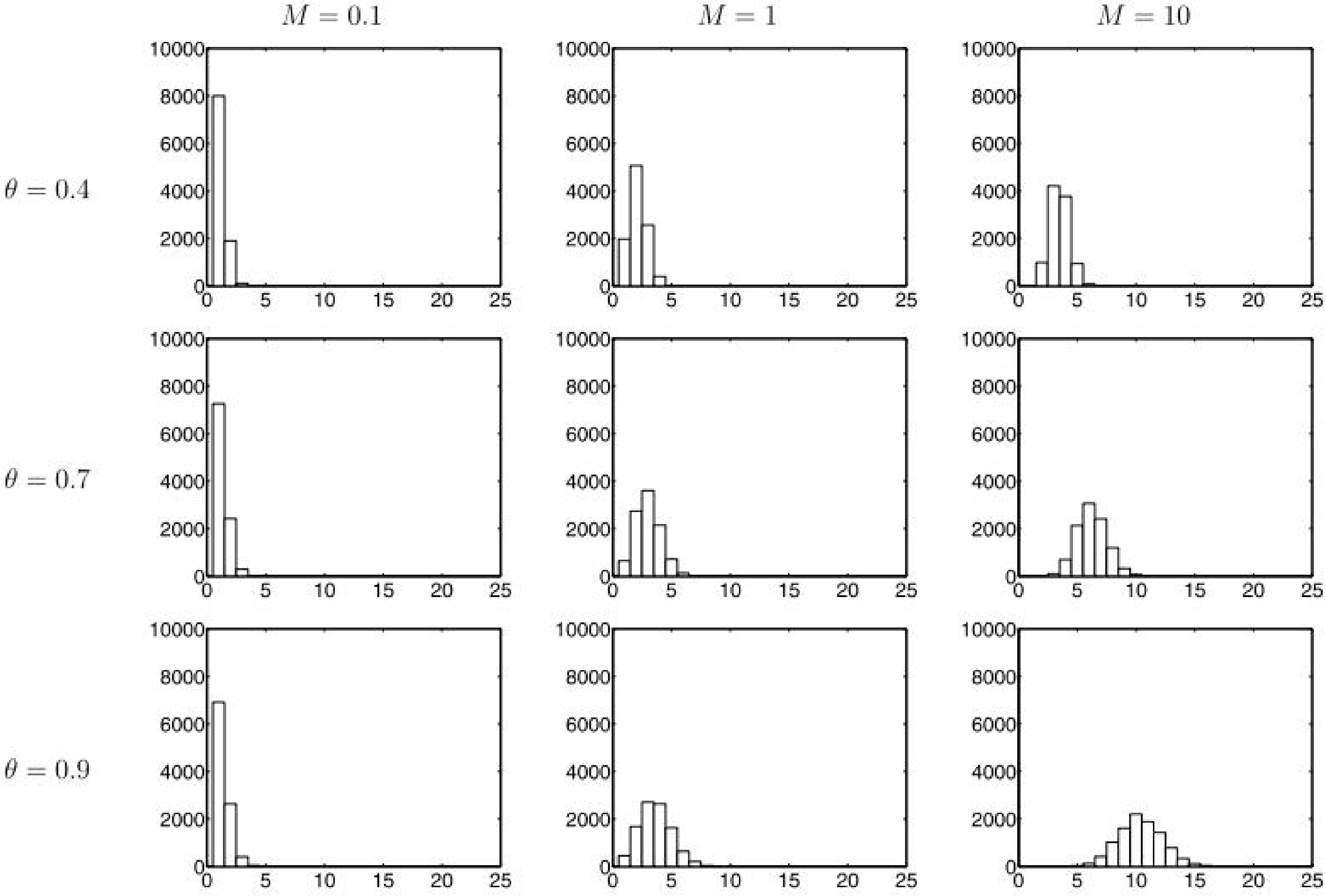
subject to: $\sum_{j=1}^{\infty} \lambda_j < +\infty$

suggested:

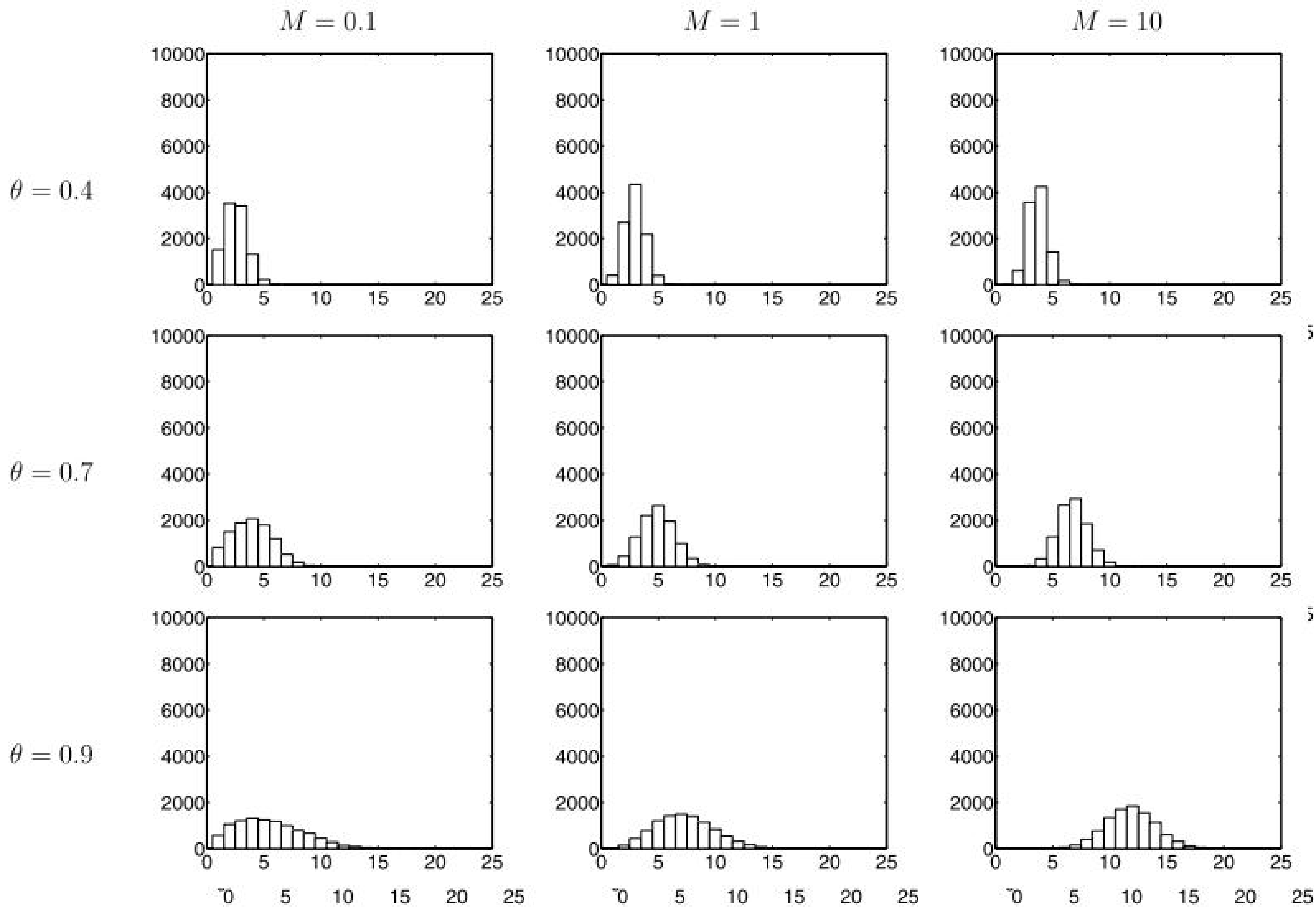
$$\mathrm{E}(\lambda_j) = M q_j$$

$$q_j = (1 - \theta)\theta^{j-1}.$$

Examples: Gamma



Examples: Inverse-Gaussian



Joint Density, now with v

$$f(y, v, u, d) = \exp(-v\Lambda)\mathbf{1}(u < \xi_d)\lambda_d/\xi_d K(y; \phi_d).$$

Clearly the marginal density is

$$f(y, d) = \frac{\lambda_d}{\Lambda} K(y; \phi_d).$$

Likelihood after n samples

$$\prod_{i=1}^{n} \exp(-v_i \Lambda) \mathbf{1}(u_i < \xi_{d_i}) \lambda_{d_i} / \xi_{d_i} K(y_i ; \phi_{d_i}).$$

$$v^{n-1} \exp(-v \Lambda) \prod_{i=1}^{n} \mathbf{1}(u_i < \xi_{d_i}) \lambda_{d_i} / \xi_{d_i} K(y_i ; \phi_{d_i}).$$

Simulating lambda's

$$\exp\{-\upsilon\Lambda_m\}\pi_m^\star(\Lambda_m)\prod_{j=1}^{m}\exp\{-\upsilon\lambda_j\}\lambda_j^{n_j}\pi_j(\lambda_j),$$

m is the # of atoms from the previous iteration

$$n_j = \sum_{i=1}^{n} \mathbf{1}(d_i = j)$$

for which $\Lambda_{m'} < \min_i\{u_i\}$

Finding weights > u_i

Need to find the smallest value of m'

for which $\Lambda_{m'} < \min_i\{u_i\}$

$[\lambda_j = x, \Lambda_j = \Lambda_{j-1} - x | \Lambda_{j-1}]$ is given by

$$f(x) \propto \pi_j(x)\pi_j^\star(\Lambda_{j-1} - x),$$

$$0 < x < \Lambda_{j-1}.$$

Dependent sampler

should this be K?

1. $\pi(\mu_j, \sigma_j^2 | \cdots) \propto p_0(\mu_j, \sigma_j^2) \prod_{d_i=j} \mathrm{N}(y_i; \mu_j, \sigma_j^2)$.
2. $\pi(\lambda_j) \propto \lambda_j^{n_j} \exp\{-v\lambda_j\}$ and $\pi(\Lambda_m) \propto \exp\{-v\Lambda\} \times \pi^\star(\Lambda_m)$.
3. $\pi(u_i | \cdots) \propto \mathbf{1}(0 < u_i < \xi_{d_i})$. tricky to calculate
4. $\mathrm{P}(d_i = k | \cdots) \propto \mathbf{1}(k : \xi_k > u_i) w_k / \xi_k \mathrm{N}(y_i; \mu_k, \sigma_k^2)$.
5. v is Gamma distributed with shape parameter $n - 1$ and mean $(n-1)/(\Lambda_m + \sum_{i=1}^m \lambda_i)$.

Condition distribution of v

$$\mathrm{E}[\exp\{-v\Lambda_m\}]v^{n-1}\exp\left\{-v\sum_{j=1}^{m}\lambda_j\right\}$$

Independent sampler

1. $\pi(\mu_j, \sigma_j^2 | \cdots) \propto p_0(\mu_j, \sigma_j^2) \prod_{d_i=j} \mathrm{N}(y_i; \mu_j, \sigma_j^2)$.
2. $\pi(\lambda_j) \propto \lambda_j^{n_j} \exp\{-v\lambda_j\}$.
3. $\pi(u_i | \cdots) \propto \mathbf{1}(0 < u_i < \xi_{d_i})$.
4. $\mathrm{P}(d_i = k | \cdots) \propto \mathbf{1}(k : \xi_k > u_i) w_k / \xi_k \mathrm{N}(y_i; \mu_k, \sigma_k^2)$.
5. $\pi(v) \propto \mathrm{E}[\exp\{-v\Lambda_m\}] v^{n-1} \exp\{-v \sum_{i=1}^m \lambda_i\}$ which is a univariate distribution and can be updated using standard methods.

Results?

Generally agrees w/ established samplers
Independent faster than dependent
Best independent slice-efficient sampler "outperforms" retrospective sampler (on autocorrelation)
Better startup time than retrospective sampler?
Apparently retrospective samplers are difficult to implement?

Slightly worse integrated autocorrelation times than retrospective sampler
Additional mystery parameter κ

Multipurpose bar graph