

```
[width=14pt,height=12pt]beamericonbookbeamericonbook  
[width=14pt,height=12pt]beamericonbookshadedbeamericonbook.20  
beamericonbook.!20opaquebeamericonbookshaded  
beamericonbook.!15opaquebeamericonbookshaded  
beamericonbook.!10opaquebeamericonbookshaded  
beamericonbook.!5opaquebeamericonbookshaded  
beamericonbook.!2opaquebeamericonbookshaded  
[width=11pt,height=14pt]beamericonarticlebeamericonarticle  
[width=11pt,height=14pt]beamericonarticleshadedbeamericonarticle.20  
beamericonarticle.!20opaquebeamericonarticleshaded  
beamericonarticle.!15opaquebeamericonarticleshaded  
beamericonarticle.!10opaquebeamericonarticleshaded  
beamericonarticle.!5opaquebeamericonarticleshaded  
beamericonarticle.!2opaquebeamericonarticleshaded  
@keypgfimagewidth @keypgfimageheight @keypgfimagepage  
@keypgfimageinterpolate[true] @keypgfimagemaskifundefinedpgf@mask@1  
@keypgfmaskmatte
```

On the flexibility of acceptance probabilities in auxiliary variable Metropolis-Hastings algorithms

On the flexibility of acceptance probabilities in auxiliary variable Metropolis-Hastings algorithms

Geir Storvik

Dept. of Mathematics, University of Oslo
SFI², Statistics for innovation
Norwegian Computing Center
SFF-CEES, Ecology and Evolution synthesis

Warwick March 16-20 2009

Outline

MC-methods in complex situations

Simulation using auxiliary variables

Common framework

Importance sampling

Metropolis-Hastings

Sequentially generated proposals

Non-sequential algorithms

Summary/discussion

Monte Carlo estimation

- ▶ Interest in $\theta = \int_y f(y)\pi(y)dy = E^\pi[f(y)]$
- ▶ Markov Chain Monte Carlo
 - ▶ Estimated through $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n f(y_i^*)$
 - ▶ y_1^*, y_2^*, \dots generated through a Markov chain with π as invariant distribution
- ▶ Importance sampling
 - ▶ Estimated through $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n \frac{\pi(y_i^*)}{q(y_i^*)} f(y_i^*)$, y_i^* iid from $q(\cdot)$
- ▶ Complex situations: Simple methods not enough
 - ▶ Multiple modes
 - ▶ Jumps between models
 - ▶ Densities with long, narrow contours

Combining importance sampling/MCMC

- ▶ MCMC gives samples close to $\pi(y)$
- ▶ Importance sampling correct for deviance
- ▶ Possible to combine?
- ▶ Using within importance sampling:
 - ▶ Generate y^* through an MCMC algorithm $x_1^*, x_2^*, \dots, x_{t+1}^* = y^*$
 - ▶ Importance weight $\pi(y^*)/q_y(y^*)$ with

$$q_y(y) = q_y(x_{t+1}) = \int_{x_{1:t}} q_1(x_1) \prod_{j=2}^{t+1} q(x_j|x_{j-1}) dx_{1:t}$$

- ▶ Problem: $q_y(y)$ not easy to compute
- ▶ Other proposal schemes of interest
- ▶ Many “proper” weights possible through extended space formulation

Use of auxiliary variables within MH

- ▶ Current value y , simulate $y^* \sim q_y(\cdot|y)$

$$y^{new} = \begin{cases} y^* & \text{with prob } \alpha(y; y^*) = \min\left\{1, \frac{\pi(y^*)q_y(y|y^*)}{\pi(y)q_y(y^*|y)}\right\} \\ y & \text{otherwise} \end{cases}$$

- ▶ Simulation through auxiliary variables

$$x_1^* \sim q_1(x_1^*|y),$$

$$x_j^* \sim q_j(x_j^*|x_{j-1}^*), j = 2, \dots, t$$

$$y^* \sim q_{y|x}(y^*|x_t^*).$$

- ▶ What acceptance probability should now be used?
- ▶ Ideal $\alpha(y; y^*)$ usually not possible to evaluate
- ▶ Many “proper” choices are available through working in extended space.

Examples from literature

- ▶ Annealed importance sampling (Neal, 2001)
- ▶ Mode jumping (Tjelmeland and Hegstad, 2001; Jennison and Sharp, 2007)
- ▶ Model selection and reversible jump MCMC (Al-Awadhi et al., 2004)
- ▶ Delayed rejection sampling (Tierney and Mira, 1999; Green and Mira, 2001)
- ▶ Pseudo-marginal algorithms (Beaumont, 2003)
- ▶ Likelihoods with intractable normalising constants (Møller et al., 2006)
- ▶ Directional MCMC (Eidsvik and Tjelmeland, 2006)
- ▶ Multi-try methods and particle proposals (Liu et al., 2000; Andrieu et al., 2008)

Common framework

- ▶ Different ways of motivating the algorithms
- ▶ Different ways of proving their validity
- ▶ Possible to put all into common framework
- ▶ “Standard” use of Imp.samp/MCMC on extended spaces
- ▶ Easier to understand
- ▶ Easier to construct alternative acceptance probabilities.
- ▶ Toolbox for constructing/proving validity of other algorithms.

Auxiliary variables and importance sampling

- ▶ Simulate y^* through $x^* \sim q_x(\cdot)$ and $y^* \sim q_{y|x}(\cdot|x^*)$.
- ▶ For any conditional distribution $h(x|y)$,

$$\theta = \int_y f(y) \pi(y) dy = \int_y \int_x f(y) \pi(y) h(x|y) dx dy$$

- ▶ $\theta = \int_y \int_x f(y) \frac{\pi(y) h(x|y)}{q_x(x) q_{y|x}(y|x)} q_x(x) q_{y|x}(y|x) dx dy$
- ▶ Importance sampling:

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n w(x^i, y^i) f(y^i)$$

$$w(x, y) = \frac{\pi(y) h(x|y)}{q_x(x) q_{y|x}(y|x)}$$

- ▶ Many possible weights for different choices of h !

Properties

- ▶ Define $q_y(y) = \int_x q_x(x)q_{y|x}(y|x)dx$. Then

$$E[w(x, y)|y] = \frac{\pi(y)}{q_y(y)}$$

$$\text{Var}[w(x, y)] \geq \text{Var}\left[\frac{\pi(y)}{q_y(y)}\right]$$

Combination of h 's

- ▶ Combination of h 's:

$$h_s(x|y), s = 1, 2, \dots \quad \text{cond. distributions}$$

imply for $a_s \geq 0, \sum_s a_s = 1$

$$h(x|y) = \sum_s a_s h_s(x|y) \quad \text{cond. distribution}$$

- ▶ Combination of w 's:

$$w_s(x, y) = \frac{\pi(y)h_s(x|y)}{q_x(x)q_{y|x}(y|x)} \quad \text{proper weight function}$$

imply

$$\sum_s a_s w_s(x, y) \quad \text{also proper weight function}$$

Proposal through MCMC-steps

- ▶ Assume $x_1^* \sim q_1(\cdot)$, $x_j^* \sim q(\cdot|x_{j-1}^*), j = 2, \dots, t+1$ with

$$\pi(x)q(y|x) = \pi(y)q(x|y)$$

- ▶ Proposal $y^* = x_{t+1}^*$
- ▶ Note: $\frac{\pi(y)}{q_y(y)} \approx 1$ for t large
- ▶ Possible to choose $h(x|y)$ such that $w(x, y) \rightarrow 1$ as $t \rightarrow \infty$?

Proposal through MCMC (cont)

- ▶ $h_s(x_{1:t}|y = x_{t+1}) = q_1(x_1) \prod_{j=2}^{s-1} q(x_j|x_{j-1}) \prod_{j=s}^t q(x_j|x_{j+1})$ imply,
 $w_s(x, y) = \frac{\pi(x_s)}{q(x_s|x_{s-1})}$
- ▶ Special cases
 - ▶ $s = 1$, $w_1(x, y) = \frac{\pi(x_1)}{q_1(x_1)}$
 - ▶ $s = t + 1$, $w_{t+1}(x, y) = \frac{\pi(x_{t+1})}{q(x_{t+1}|x_t)}$
 - ▶ Combination: $\bar{w}(x, y) = \frac{1}{t} \sum_{j=2}^{t+1} \frac{\pi(x_s)}{q(x_s|x_{s-1})}$
 - ▶ Note: $\bar{w}(x, y) \rightarrow E[w(x, y)] = 1$ as $t \rightarrow \infty$

MH and auxiliary variables

- ▶ Ideas from importance sampling can be transferred to MCMC
- ▶ Current value y , simulate $x^* \sim q_x(\cdot|y)$, $y^* \sim q_{y|x}(\cdot|x^*)$
- ▶ MH algorithm constructed in extended space (x, y) .
- ▶ $r(x, y; x^*, y^*) = \frac{w(y; x^*, y^*)}{w(y^*; x, y)}$
- ▶ Different strategies possible
 - ▶ Store previous x
 - ▶ Generate x^*, y^* and x

Proposal through “inner” MCMC steps

- ▶ Assume $x_1 \sim q_1(\cdot | y)$, $x_j^* \sim q(\cdot | x_{j-1}^*), j = 2, 3, \dots, t+1$ with

$$\pi(x)q(y|x) = \pi(y)q(x|y)$$

- ▶ Proposal $y^* = x_{t+1}^*$

$$r(y; x^*, y^*, y) = \frac{w(y; x^*, y^*)}{w(y; x^*, y)}$$

- ▶ Special cases

- ▶ $w_1(x, y) = \frac{\pi(x_1)}{q_1(x_1|y)}$

- ▶ $w_{t+1}(x, y) = \frac{\pi(x_{t+1})}{q(x_{t+1}|x_t)}$

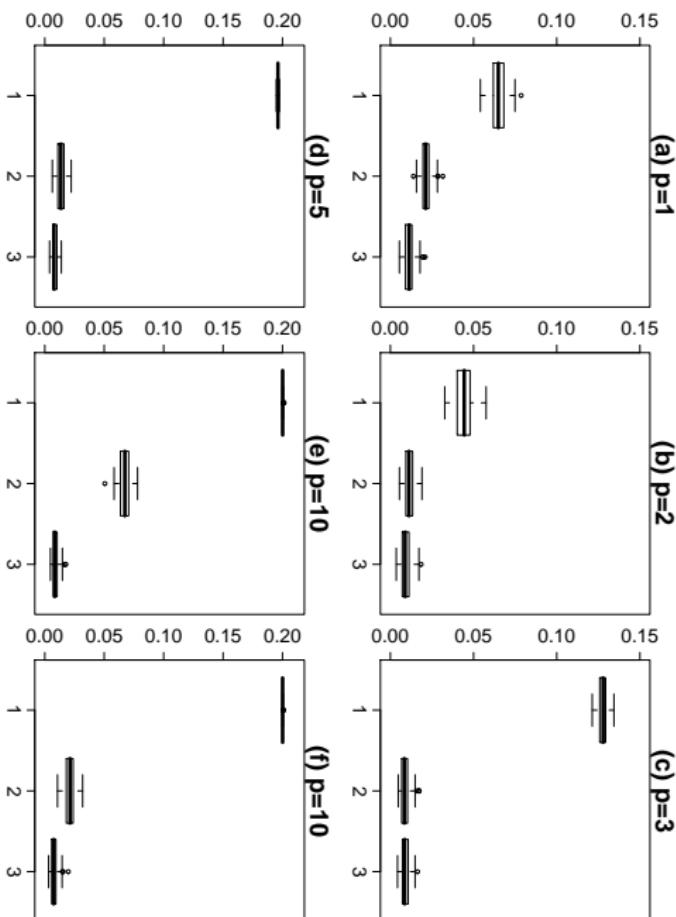
- ▶ $\bar{w}(x, y) = \frac{1}{t} \sum_{j=2}^{t+1} \frac{\pi(x_s)}{q(x_s|x_{s-1})}$

- ▶ Note: $\bar{w}(x, y) \rightarrow E[w(x, y)] = 1$ as $t \rightarrow \infty$

Example

- ▶ $\pi(y) = \beta N(y; \mu_1, I) + (1 - \beta)N(y; \mu_2, I)$ where
 $\mu_{1,1} = -\mu_{2,1} = -10$ while $\mu_{i,j} = 0, i = 1, 2, j = 2, \dots, p.$
- ▶ $x_1^* \sim N(0, \sigma_{large}^2 I)$
- ▶ $t + 1 = 10$ “inner” MH steps (discrete Langevin diffusion)
- ▶ Previous x stored
- ▶ $M = 20$ outer MCMC-steps.
- ▶ $w_1(x, y) = \frac{\pi(x_1)}{q_1(x_1)}, w_{t+1}(x, y) = \frac{\pi(x_{t+1})}{q(x_{t+1}|x_t)}, \bar{w}(x, y)$

Dist to truth for different weight functions



More general schemes

- ▶ $x^* \sim q_x(\cdot|y)$, $y^* \sim q_{y|x}(\cdot|x^*)$.
- ▶ Some sequential structure in generation of $x^* = (x_1^*, \dots, x_t^*)$.
- ▶ Examples
 - ▶ Annealed importance sampling (Neal, 2001)
 - ▶ Mode jumping (Tjelmeland and Hegstad, 2001; Jennison and Sharp, 2007)
 - ▶ Model selection and reversible jump MCMC (Al-Awadhi et al., 2004)
 - ▶ Particle proposals (Andrieu et al., 2008)

RJMCMC proposals (Al-Awadhi et al., 2004)

- ▶ x_1^* : Jump between models
- ▶ $x_j^*, j = 2, \dots, t+1$: Jump within model
- ▶ (Al-Awadhi et al., 2004): $w_1(y; x^*, y^*) = \frac{\pi(n)\pi(x_1^*)}{q_M(n|m, y)q_n(x_1^*|y)}$
- ▶ Alternatives

$$w_s(y; x^*, y^*) = \frac{\pi(n)\pi(x_s^*)}{q_M(n|m, y)q_n(x_s^*|x_{s-1}^*)}$$
$$r = \frac{\pi(n)q_M(m|n, y^*) \sum_{s=2}^{t+1} \frac{\pi_n(x_s^*)}{q_n(x_s^*|x_{s-1}^*)}}{\pi(m)q_M(n|m, y) \sum_{s=2}^{t+1} \frac{\pi_m(x_s)}{q_m(x_s|x_{s-1})}}.$$

- ▶ Converges towards $\pi(n)q_M(m|n, y^*)/\pi(m)q_M(n|m, y)$

Particle proposals

- ▶ Generate parallel sequences $\{x_{j,1:t+1}^*, j = 1, \dots, N\}$
- ▶ Choose $K \in \{1, \dots, N\}$, put $y^* = x_{K,t+1}^*$.
- ▶ Each sequence independently: $x_{j,i}^* \sim q_i(x_{j,i}^* | x_{j,i-1}^*)$

$$\Pr(K) \propto \frac{\pi(x_{K,t+1}^*)}{q_{t+1}(x_{K,t+1}^* | x_{K,t}^*)}.$$

- ▶ $h(x|y, x^*, y^*) = N^{-1} \delta(x_{K,t+1} - y) \prod_{i=1}^t q_i(x_{K,i} | x_{K,i-1}) \prod_{j \neq K} \prod_{i=1}^{t+1} q_i(x_{j,i} | x_{j,i-1})$ gives

$$r(y; x^*, y^*, x) = \frac{\sum_{j=1}^N \frac{\pi(x_{j,t+1}^*)}{q_{t+1}(x_{j,t+1}^* | x_{j,t}^*)}}{\sum_{j=1}^N \frac{\pi(x_{j,t+1})}{q_{t+1}(x_{j,t+1} | x_{j,t})}}.$$

Non-sequential examples

- ▶ Pseudo-marginal algorithms (Beaumont, 2003; Andrieu and Roberts, 2009)
- ▶ Likelihoods with intractable normalising constants (Møller et al., 2006)
- ▶ Delayed rejection sampling (Tierney and Mira, 1999; Green and Mira, 2001)

Delayed rejection sampling

- ▶ $x_1^* \sim q_1(\cdot|y)$, accepted with prob. $\alpha_1(y; x_1^*)$
- ▶ If rejected, $x_2^* \sim q_2(\cdot|x_1^*, y)$, accepted with prob. $\alpha_2(y; x_1^*, x_2^*)$
- ▶ $y^* = x_1^*$ with prob α_1 , otherwise $= x_2^*$.
- ▶ Tierney and Mira (1999)

$$\alpha_2(y; x_1^*, x_2^*) = \min \left\{ 1, \frac{\pi(x_2^*) q_1(x_1^*|x_2^*) q_2(y|x_1^*, x_2^*) [1 - \alpha_1(x_2^*; x_1^*)]}{\pi(y) q_1(x_1^*|y) q_2(x_2^*|x_1^*, y) [1 - \alpha_1(y; x_1^*)]} \right\}$$

- ▶ Alternative: Generate $x_1 \sim q_1(x_1|y^*)$,

$$\alpha_2(y; x_1^*, x_2^*) = \min \left\{ 1, \frac{\pi(y^*) q_2(y|x_1, y^*) [1 - \alpha_1(y^*; x_1)]}{\pi(y) q_2(y^*|x_1^*, y) [1 - \alpha_1(y; x_1^*)]} \right\}$$

- ▶ $q_2 = \pi \Rightarrow \alpha_2(y; x_1^*, x_2^*) = \min \left\{ 1, \frac{1 - \alpha_1(y^*; x_1)}{1 - \alpha_1(y; x_1^*)} \right\}$

Example

$$z_i = \mu + \eta_i + \varepsilon_i,$$

$$i = 1, \dots, n$$

$$\varepsilon \sim N(0, \tau_1^{-1}),$$

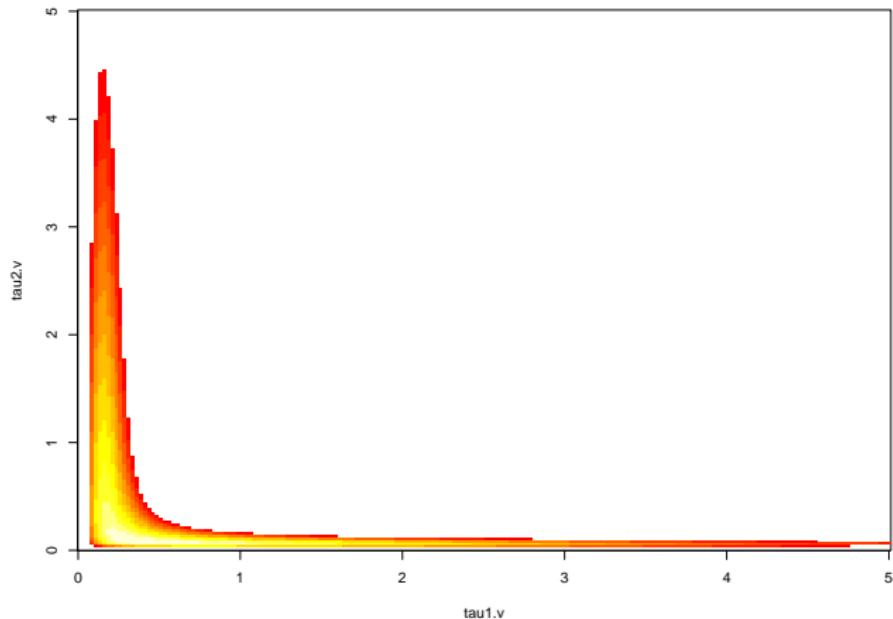
iid

$$\eta_i | \eta_{-i} \sim N(\beta n_i^{-1} \sum_{j \sim i} \eta_j, n_i^{-1} \tau_2^{-1}),$$

CAR

Interest in posterior for (τ_1, τ_2) .

Posterior for (τ_1, τ_2)



Delayed rejection sampling

- ▶ First stage:

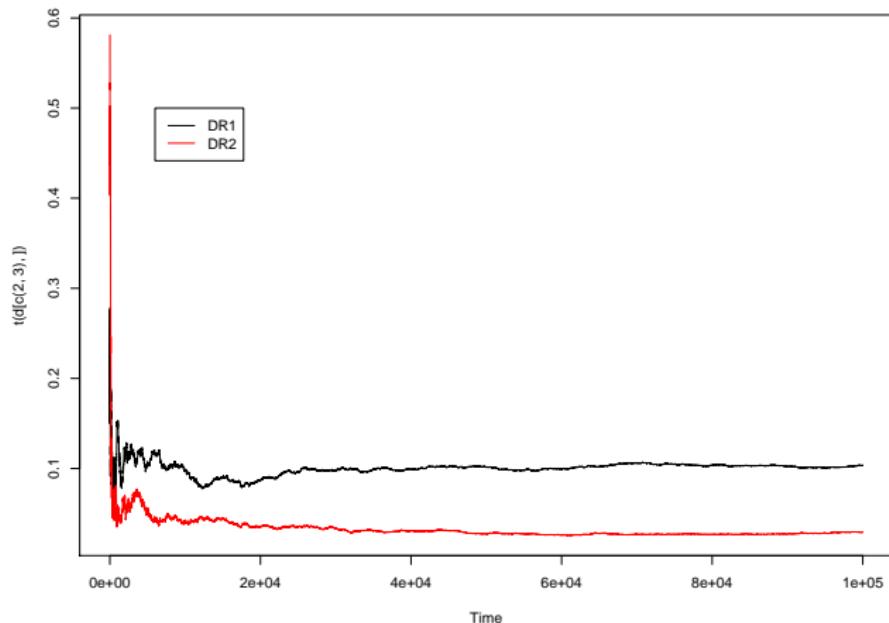
$$\tau_j^1 = \tau_j f_j, \quad q(f_j) = (1 + f^{-1}) I(f_j \in [F^{-1}, F]), j = 1, 2$$

- ▶ Second stage: Re-parametrisation

$$\begin{aligned}\tau &= \frac{\tau_1 \tau_2}{\tau_1 + \tau_2}, & r &= \frac{\tau_2}{\tau_1 + \tau_2} \\ r &\sim [0, 1], & \tau &\sim \pi(\tau|r)\end{aligned}$$

- ▶ Compare two choices for α_2 .

KS-distance to posterior for (τ_1, τ_2)



Pseudo-marginal algorithms

- ▶ Target given through $\pi(y) = \int_x \bar{\pi}(x, y) dx$
- ▶ Simulate $y^* \sim q_y(y^*|y)$
- ▶ Simulate $x_1^*, \dots, x_t^* \sim q_{x|y}(x^*|y^*)$
- ▶ $h_s(x_{1:N}|y) = \bar{\pi}(x_s|y) \prod_{j \neq s} q_{x|y}(x_j|y)$
- ▶ Acceptance ratio (Beaumont, 2003)

$$r_{x,y}(x^*, y^*) = \frac{w(y, x^*, y^*)}{w(y^*, x, y)}$$

where

$$w(y, x^*, y^*) = \frac{1}{t} \sum_{i=1}^t \frac{\bar{\pi}(x_i^*, y^*)}{q_{x|y}(x_i^*|y^*) q_y(y^*|y)} \approx \frac{\pi(y^*)}{q_y(y^*|y)}$$

- ▶ Variant: Generate new x 's each time

Intractable normalising constants (Møller et al., 2006)

$$\begin{aligned}\pi(y) &= p(y|z) = C^{-1} p(y)p(z|y) \\ p(z|y) &= Z^{-1}(y)\tilde{p}(z|y)\end{aligned}$$

- ▶ (x, y) is the current state
- ▶ $q(x^*, y^*|x, y) = q_y(y^*|x, y)q_{x|y}(x^*|x, y, y^*)$ with
 $q_{x|y}(x^*|x, y, y^*) = Z^{-1}(y^*)\tilde{p}(x^*|y^*)$
- ▶ $h(x|y)$ arbitrary, but state space similar to z .

$$r(x, y; x^*, y^*) = \frac{p(y^*)\tilde{p}(z|y^*)h(x^*|y^*)q_y(y|x^*, y^*)\tilde{p}(x|y)}{p(y)\tilde{p}(z|y)h(x|y)q_y(y^*|x, y)\tilde{p}(x^*|y^*)}$$

- ▶ $h(x|y, x^*, y^*) = \delta(x - x^*)$ gives
- $$r(x^*, y^*|x, y) = \frac{p(y^*)\tilde{p}(z|y^*)q_{y|x}(y^*|y)\tilde{p}(x^*|y)}{p(y)\tilde{p}(z|y)q_{y|x}(y|y^*)\tilde{p}(x^*|y^*)}$$

Summary/discussion

- ▶ General framework: $x^* \sim q_x(\cdot|y)$, $y^* \sim q_{y|x}(\cdot|x^*)$
- ▶ Many existing algorithms within this framework
- ▶ Many different weight functions/acceptance probabilities are possible
- ▶ General framework:
 - ▶ Tool for constructing new algorithm
 - ▶ Tool for constructing alternative weight functions
 - ▶ Common understanding of many different algorithms
- ▶ Further work: Experimental/theoretical results on properties of different weight functions

References I

- Al-Awadhi, F., M. Hurn, and C. Jennison (2004). Improving the acceptance rate of reversible jump mcmc proposals. *Statistics & Probability Letters* 69, 189–198.
- Andrieu, C., A. Doucet, and R. Holenstein (2008). Particle Markov chain Monte Carlo. Preprint.
- Andrieu, C. and G. O. Roberts (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Annals of Statistics*. To be published.
- Beaumont, M. (2003). Estimation of Population Growth or Decline in Genetically Monitored Populations. *Genetics* 164(3), 1139–1160.
- Eidsvik, J. and H. Tjelmeland (2006). On directional metropolis-hastings algorithms. *Stat. Comput* 16, 93–106.
- Green, P. and A. Mira (2001). Delayed rejection in reversible jump Metropolis-Hastings. *Biometrika* 88(4), 1035–1053.
- Jennison, C. and R. Sharp (2007). Mode jumping in MCMC: Adapting proposals to the local environment. Talk at Conference to honour Allan Seheult, Durham, March 2007. Available at <http://people.bath.ac.uk/mascj/>.
- Liu, J., F. Liang, and W. Wong (2000). The multiple-try method and local optimization in Metropolis sampling. *Journal of the American Statistical Association* 95(449), 121–134.
- Møller, J., A. Pettitt, R. Reeves, and K. Berthelsen (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika* 93(2), 451–458.
- Neal, R. M. (2001). Annealing importance sampling. *Statistics and Computing* 2, 125–139.
- Tierney, L. and A. Mira (1999). Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine* 18(1718), 2507–2515.
- Tjelmeland, H. and B. Hegstad (2001). Mode jumping proposals in MCMC. *Scand. J. Statist.* 28(1), 205–223.