CMPE 58N - Lecture 7.
Monte Carlo methods

Reversible Jump MCMC

Department of Computer Engineering,

Boğaziçi University, Istanbul, Turkey

Instructor: A. Taylan Cemgil

Spring 2009

Model Selection

- Number of things you don't know is one of the things you don't know

“Model Selection” Example

Given an unknown number of fair dice with outcomes
$\lambda_1, \lambda_2, \ldots, \lambda_n$,

$$\mathcal{D} = \sum_{i=1}^{n} \lambda_i$$

How many dice are there when $\mathcal{D} = 9$?
Assume that any number n is equally likely *a-priori*

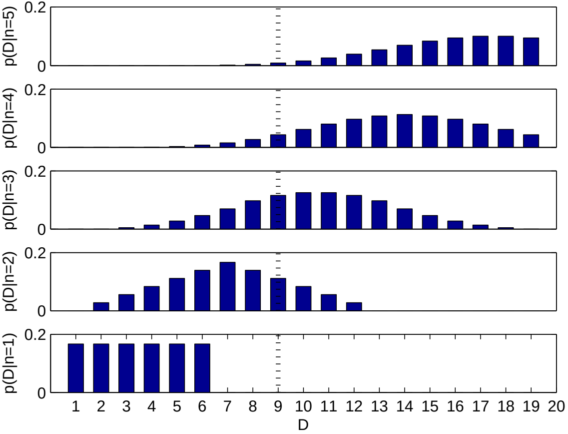
“Model Selection” Example

Given all n are equally likely (i.e., $p(n)$ is flat), we calculate (formally)

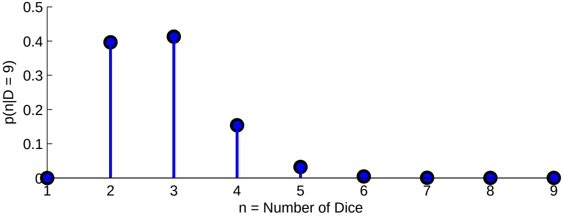
$$p(n|\mathcal{D}=9) = \frac{p(\mathcal{D}=9|n)p(n)}{p(\mathcal{D})} \propto p(\mathcal{D}=9|n)$$

$$\begin{aligned}
p(\mathcal{D}|n=1) &= \sum_{\lambda_1} p(\mathcal{D}|\lambda_1)p(\lambda_1) \\
p(\mathcal{D}|n=2) &= \sum_{\lambda_1}\sum_{\lambda_2} p(\mathcal{D}|\lambda_1,\lambda_2)p(\lambda_1)p(\lambda_2) \\
\dots \\
p(\mathcal{D}|n=n') &= \sum_{\lambda_1,\dots,\lambda_{n'}} p(\mathcal{D}|\lambda_1,\dots,\lambda_{n'}) \prod_{i=1}^{n'} p(\lambda_i)
\end{aligned}$$

$$p(\mathcal{D}|n) = \sum_{\boldsymbol{\lambda}} p(\mathcal{D}|\boldsymbol{\lambda}, n) p(\boldsymbol{\lambda}|n)$$

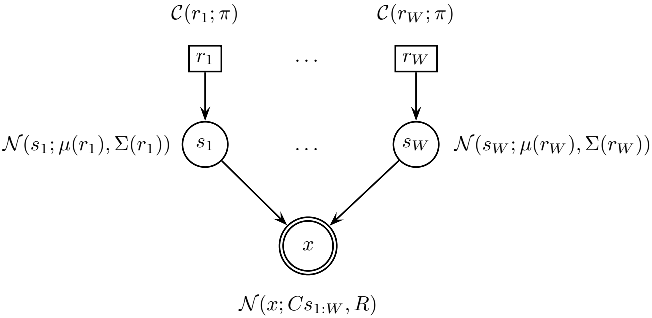


"Model Selection" Example



- Complex models are more flexible but they spread their probability mass
- Bayesian inference inherently prefers "simpler models" – Occam's razor
- Computational burden: We need to sum over all parameters λ

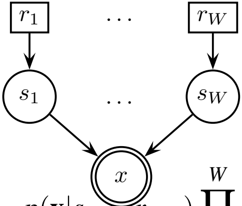
Bayesian Variable Selection



- Generalized Linear Model – Column's of C are the basis vectors
- The exact posterior is a mixture of 2^W Gaussians
- When W is large, computation of posterior features becomes intractable.

Generative model

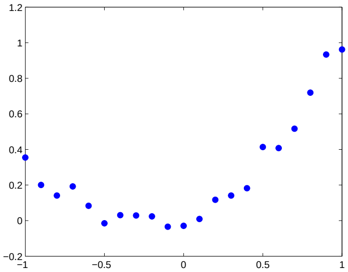
$$
\begin{aligned}
r_i &\sim \mathcal{C}(r_i; \pi) \\
s_i | r_i &\sim \mathcal{N}(s_i; \mu(r_i), \Sigma(r_i)) \\
\mathbf{x} | s_{1:W} &\sim \mathcal{N}(\mathbf{x}; C s_{1:W}, R) \\
C &\equiv \left[\begin{array}{ccccc} C_1 & \dots & C_i & \dots & C_W \end{array} \right]
\end{aligned}
$$



$$
p(\mathbf{x}, s_{1:W}, r_{1:W}) = p(\mathbf{x} | s_{1:W}, r_{1:W}) \prod_{i=1}^{W} p(s_i | r_i) p(r_i)
$$

Example 1: Variable selection in Polynomial Regression

Given $\{t_j, x(t_j)\}_{j=1\ldots J}$, what is the order N of the polynomial?



$$x(t) = \sum_{i=0}^{N} s_{i+1} t^i + \epsilon(t)$$

Ex1: Regression

$$\begin{aligned} \mathbf{t} &= \begin{pmatrix} t_1 & t_2 & \dots & t_J \end{pmatrix}^\top \\ C &\equiv \begin{pmatrix} \mathbf{t}^0 & \mathbf{t}^1 & \dots & \mathbf{t}^{W-1} \end{pmatrix} \end{aligned}$$

```
>> C = fliplr(vander(0:4))  % Van der Monde matrix
     1     0     0     0     0
     1     1     1     1     1
     1     2     4     8    16
     1     3     9    27    81
     1     4    16    64   256
```

$$\begin{aligned} r_i &\sim \mathcal{C}(r_i; 0.5, 0.5) \qquad\qquad r_i \in \{\text{on}, \text{off}\} \\ s_i | r_i &\sim \mathcal{N}(s_i; 0, \Sigma(r_i)) \\ \mathbf{x} | s_{1:W} &\sim \mathcal{N}(\mathbf{x}; C s_{1:W}, R) \end{aligned}$$

$$\Sigma(r_i = \text{on}) \gg \Sigma(r_i = \text{off})$$

Ex1: Regression

To find the “active” basis functions we need to calculate

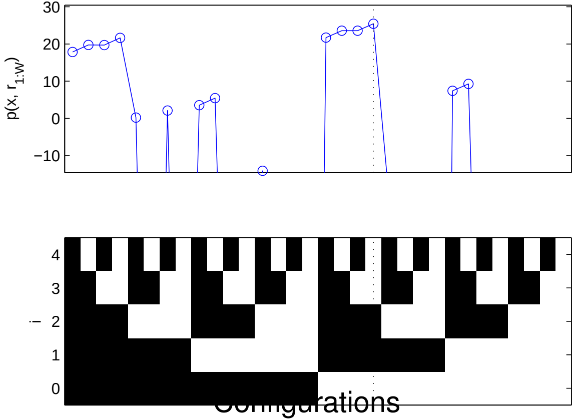
$$r^*_{1:W} \equiv \operatorname*{argmax}_{r_{1:W}} p(r_{1:W}|\mathbf{x}) \quad = \quad \operatorname*{argmax}_{r_{1:W}} \int ds_{1:W} p(\mathbf{x}|s_{1:W}) p(s_{1:W}|r_{1:W}) p(r_{1:W})$$

Then, the reconstruction is given by

$$\begin{aligned} \hat{x}(t) &= \left\langle \sum_{i=0}^{W-1} s_{i+1} t^i \right\rangle_{p(s_{1:W}|\mathbf{x}, r^*_{1:W})} \\ &= \sum_{i=0}^{W-1} \langle s_{i+1} \rangle_{p(s_{i+1}|\mathbf{x}, r^*_{1:W})} t^i \end{aligned}$$

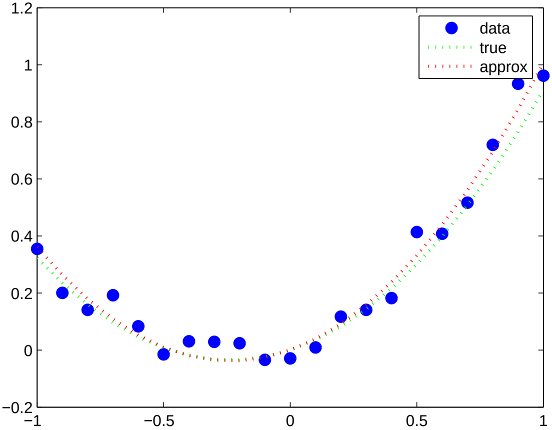
Ex1: Regression

All on



All off

Ex1: Regression



Model Selection

- Number of things you don't know is one of the things you don't know
- k: Number of parameters
- $\theta_k \in \mathbb{R}^{n_k}$, n_k is model dimension

Model Selection via Reversible Jump

- "State Space"

$$(k, \theta_k) \in \bigcup_{k \in \mathcal{K}} \{\{k\} \times \mathbb{R}^{n_k}\}$$

- Reversible Jump $=$ Metropolis-Hastings on this general and "non-standard" state space
- We use Metropolis-Hastings to build a suitable reversible chain.
-

Metropolis-Hastings

- The target distribution is $\pi(s)$.
- We choose a suitable proposal distribution $q(s'|s)$
- We define the *acceptance probability* of a jump from s to s' as

$$a(s \to s') \equiv \min\{1, \frac{\pi(s')q(s|s')}{\pi(s)q(s'|s)}\}$$

- We have verified the detailed balance for the MH transition Kernel T

Metropolis-Hastings Transition Kernel

- Transition Kernel

$$T(s'|s) = \underbrace{q(s'|s)a(s \to s')}_{\text{Accept}} + \underbrace{\delta(s'-s)\int ds' q(s'|s)(1-a(s \to s'))}_{\text{Reject}}$$

- We know that T satisfies detailed balance, i.e.,

$$T(s|s')\pi(s') = T(s'|s)\pi(s)$$

Implications of Detailed Balance

- Consider now the probability that the chain is in $s \in S$ and jumps to $s' \in S'$

$$\Pr\{S \to S'\} \equiv \int_{(s,s') \in S \times S'} ds ds' T(s'|s) \pi(s)$$

- Detailed balance says simply that this probability is equal to the probability that the chain is in $s \in S$ and jumps to $s' \in S'$

$$\Pr\{S' \to S\} \equiv \int_{(s,s') \in S \times S'} ds ds' T(s|s') \pi(s')$$

- (Technicality: S and S' need to be "measurable")

Implications of Detailed Balance

- On $S \bigcap S'$, i.e., when the chain is not moving with $s' = s$, we have

$$\int_{(s,s')\in S\times S'} ds ds' \delta(s'-s) \int ds' q(s'|s)(1-a(s\to s')) = \int_{(s,s')\in S\times S'} ds ds' \delta(s'-s) \int ds q(s|s')(1-a(s'\to s))$$

- This implies that ...

Equilibrium Probability of a Jump

$$\int_{(s,s')\in S\times S'} ds ds' q(s'|s)\pi(s)a(s\to s') =$$
$$\int_{(s,s')\in S\times S'} ds ds' q(s|s')\pi(s')a(s'\to s)$$

Metropolis-Hastings

- Consider how we implement MH in an abstract sense
  - We draw some random numbers $u \sim g(u)$
  - The new state is a deterministic function of the current state s and u, i.e.,

$$s' = h(s, u)$$

  - The reverse jump would be given by new random numbers $u' \sim g'(u')$

$$s = h'(s', u')$$

Equilibrium Probability of a Jump

- In other words, we implement the proposal as

$$
\begin{aligned}
q(s'|s) &= \int du \delta(s' - h(s,u)) g(u) \\
&= \int_{(s,s'=h(s,u))} du g(u)
\end{aligned}
$$

where $g(u)$ is the joint density of u.

Equilibrium Probability of a Jump

- The probability jumping from S to S' is

$$\Pr\{S \to S'\} = \int_{(s,s'=h(s,u))\in S\times S'} ds du g(u) \pi(s) a(s \to s')$$

Equilibrium Probability of a Jump

- The probability of jumping from S' to S is

$$\Pr\{S' \to S\} = \int_{(s=h'(s',u'),s')\in S\times S'} ds' du' g(u') \pi(s') a(s' \to s)$$

- By detailed balance, we have

$$\Pr\{S \to S'\} = \Pr\{S' \to S\}$$

- Detailed balance holds iff ...

Transformation $(s, u) \rightarrow (s', u')$

$$g(u)\pi(s)a(s \rightarrow s') = g(u')\pi(s')a(s' \rightarrow s) \left| \frac{\partial(s', u')}{\partial(s, u)} \right|$$

$$a(s \rightarrow s') = \min\left\{1, \frac{g(u')\pi(s')}{g(u)\pi(s)} \left| \frac{\partial(s', u')}{\partial(s, u)} \right| \right\}$$

- Provided that $s'(s, u)$ and $u'(s, u)$ (and their inverses) are differentiable, so that the Jacobian exists

Dimension Matching

- We need

$$
\begin{aligned}
\begin{pmatrix} s \\ u \end{pmatrix} &\in \mathbb{R}^{n+m} \\
\begin{pmatrix} s' \\ u' \end{pmatrix} &\in \mathbb{R}^{n'+m'} \\
n+m &= n'+m'
\end{aligned}
$$

- Need this to be able to define the Jacobian
- Necessary but not sufficient

Toy Example (from Green, 2002)

- The state space of the chain

$$s \in \mathbb{R}^1 \cup \mathbb{R}^2$$

- Probabilities

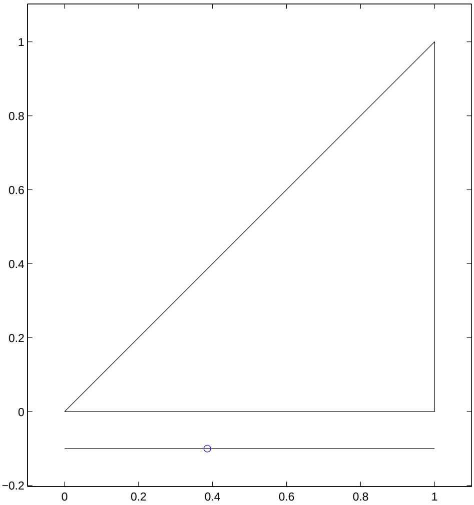
$$\begin{aligned} \Pr\{s \in \mathbb{R}\} &= p_1 \\ \Pr\{s \in \mathbb{R}^2\} &= p_2 \end{aligned}$$

- If $s \in \mathbb{R}$

$$s \sim \mathcal{U}([0, 1])$$

- If $s = (s_1, s_2) \in \mathbb{R}^2$

$$s \sim \mathcal{U}(\{s : 0 < s_2 < s_1 < 1\})$$



1
0.8
0.6
0.4
0.2
0
−0.2
0
0.2
0.4
0.6
0.8
1

Possible Moves

$(1 \leftrightarrow 1)$ In $\mathbb{R}$

$$s' \sim \mathcal{U}(s - \epsilon, s + \epsilon)$$

$(2 \leftrightarrow 2)$ In $\mathbb{R}^2$

$$(s_1, s_2)' \leftarrow (1 - s_2, 1 - s_1)$$

$(1 \leftrightarrow 2)$ Transdimensional move:

$$\mathbb{R} \rightarrow \mathbb{R}^2$$

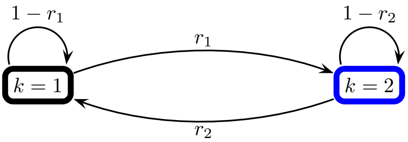
$$u \sim \mathcal{U}([0, 1])$$

$$(s'_1, s'_2) \leftarrow (s, u)$$

$$\mathbb{R}^2 \rightarrow \mathbb{R}$$

$$s \leftarrow s'_1$$

Moves



Designing the MH Kernel

- For each move $m = (k \to k')$, design a proposal distribution

$$q_m(s'|s) \equiv q(s'|s, m)q(m|s)$$

  - $j_m(s) \equiv q(m|s)$ is the probability taking the jump m when at s
  - Regular moves $(k \to k)$, no dimensionality change
  - Transdimensional moves $(k \to k')$ $k \neq k'$, need dimension matching
- Calculate the acceptance probability α_m
  - Each move must be reversible!
- The accept part of the transition Kernel is

$$\sum_m \alpha_m(s \to s')q_m(s'|s)$$

Toy Example, Regular moves

- $\mathbb{R} \to \mathbb{R}$ with $j_{1\to 1} = 1 - r_1$

$$s' \sim \mathcal{U}(s - \epsilon, s + \epsilon) = q(s'|s, m = (1 \to 1))$$

- Acceptance probability

$$\begin{aligned} \alpha_{1\to 1}(s \to s') &= \min\left\{1, \frac{q_{1\to 1}(s|s')\pi(s')}{q_{1\to 1}(s'|s)\pi(s)}\right\} \\ &= \min\left\{1, \frac{[0 < s' < 1]}{1}\right\} \end{aligned}$$

Toy Example, Regular moves

- $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ with $j_{2\rightarrow 2} = 1 - r_2$

  $$(s'_1, s'_2) \sim \delta(s'_1 - (1 - s_2))\delta(s'_2 - (1 - s_1)) = q(s'|s, m = (2 \rightarrow 2))$$

- Acceptance probability

$$\begin{aligned}
\alpha_{2\rightarrow 2}(s \rightarrow s') &= \min\left\{1, \frac{q_{2\rightarrow 2}(s|s')\pi(s')}{q_{2\rightarrow 2}(s'|s)\pi(s)}\right\} \\
&= \min\{1, 1\} = 1
\end{aligned}$$

Toy Example, Transdimensional moves

- $\mathbb{R} \to \mathbb{R}^2$ $j_{1\to 2} = r_1$

$$\begin{aligned} s'_1 &= s_1 \\ s'_2 &= u \end{aligned}$$

- $\mathbb{R}^2 \to \mathbb{R}$ $j_{2\to 1} = r_2$

$$s_1 = s'_1$$

- The Jacobian

$$\left|\frac{\partial(s'_1, s'_2)}{\partial(s_1, u)}\right| = \left|\begin{array}{cc} \partial s'_1/\partial s_1 & \partial s'_1/\partial u \\ \partial s'_2/\partial s_1 & \partial s'_2/\partial u \end{array}\right| = \left|\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array}\right| = 1$$

Toy Example, Transdimensional moves

- Proposals

$$
\begin{aligned}
g_{1\to 2}(u) &= \mathcal{U}(0,1) \\
g_{2\to 1}(u) &= 1
\end{aligned}
$$

- Remember that we implement $q(s'|s,m)$ using $g_m(u)$
- We denote the reverse move of m by $-m$
- The proposal corresponding to the reverse move of m is denoted as g_{-m} so for example

$$
g_{-(1\to 2)}(u') = g_{2\to 1}(u')
$$

Acceptance probabilities

$$\alpha_m(s \to s') = \min\left\{1, \frac{g_{-m}(u')}{g_m(u)} \frac{j_{-m}(s')}{j_m(s)} \frac{\pi(s')}{\pi(s)} \left|\frac{\partial(s', u')}{\partial(s, u)}\right|\right\}$$

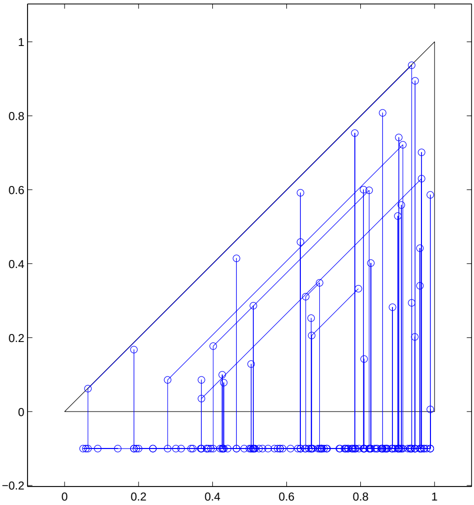
- Acceptance probability for $m = (1 \to 2)$

$$\begin{aligned}\alpha_{1\to 2}(s \to s') &= \min\left\{1, \frac{1}{1}\frac{r_2}{r_1}\frac{2\,[s'_2 < s'_1]\,p_2}{1 p_1}\,|1|\right\} \\ &= \min\left\{1, \frac{r_2}{r_1}\frac{2p_2}{p_1}\right\}[u < s_1]\end{aligned}$$

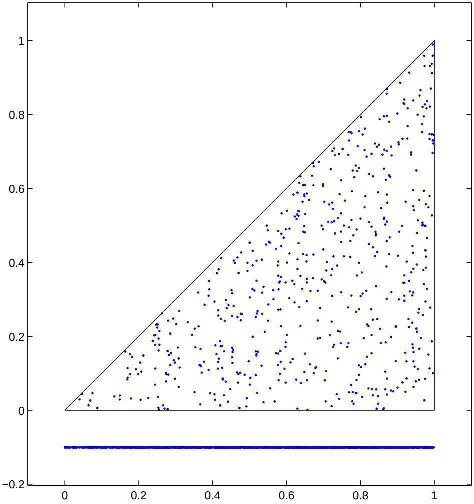
- Acceptance probability for $m = (2 \to 1)$

$$\alpha_{2\to 1}(s \to s') = \min\left\{1, \frac{1}{1}\frac{r_1}{r_2}\frac{1 p_1}{2 p_2}\,|1|\right\} = \min\left\{1, \frac{r_1}{r_2}\frac{p_1}{2p_2}\right\}$$

```
% Probability x \in R and x \in R^2
p = [0.6 0.4];
% Probability of cross dimensional move
r = [0.3 0.9];
% Proposal width in R
epsilon = 0.2;
x = 0.2; k = 1;
```



1
0.8
0.6
0.4
0.2
0
−0.2
0
0.2
0.4
0.6
0.8
1


1
0.8
0.6
0.4
0.2
0
−0.2
0
0.2
0.4
0.6
0.8
1

```
% Prior Probability x \in R and x \in R^2
p = [0.6 0.4];

% Probability of cross dimensional move
r = [0.3 0.9];
% Proposal width in R
epsilon = 0.2;

T = 5000;
% initial state
x = 0.2; k = 1;

K = zeros(1, T); X = -0.1*ones(2, T);
```

```
for e=1:T,
 if k==1, % in R
 % Which move to choose
  if rand<r(k),       % Transdimensional move
    u = rand;
    x_new = [x u];
    alpha = min(1, 2*p(2)*r(2)/(p(1)*r(1)) )...
           *double(u < x);
  else
    x_new = 2*(rand-0.5)*epsilon + x;
    alpha = min(1,  double(( 0 < x_new & x_new < 1 )) );
  end;
 elseif k==2, % in R^2
  % Which move to choose
  if rand<r(k),       % Transdimensional move
    x_new = x(1);
    alpha = min(1, p(1)*r(1)/(2*p(2)*r(2)));
  else
    x_new = 1-x(2:-1:1);
    alpha = 1;
  end;
 end;
```

```
 if rand<alpha,
  x = x_new;
  k = length(x);
 end;
 X(1:k, e) = x';
 K(e) = k;
end;
```

Observations

- Many factors that determine the efficiency
  - The jump probabilities j_m
  - Proposals q_m
  - Too many choices: To design good moves is an art

Model Selection

- Setup a MH chain on the space

$$(k, \theta_k) \in \bigcup_{k \in \mathcal{K}} \{\{k\} \times \mathbb{R}^{n_k}\}$$

- The invariant target distribution is

$$p(k, \theta_k | Y)$$

- We need typically multiple types of moves to traverse the whole space $\mathcal{X}$.
- Each move leaves π invariant but we need different move types for ergodicity

Detailed Balance for Model Choice

- Detailed balance

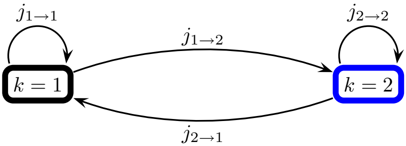
$$\int_{(s,s')\in S\times S'} ds ds' \pi(s) q_m(s'|s) \alpha_m(s \to s') = \int_{(s,s')\in S\times S'} ds ds' \pi(s') q_m(s|s') \alpha_m(s' \to s)$$

- For each m, $q_m(s'|s)$ is the *joint* distribution of move type and destination
- Acceptance probability

$$\alpha_m(s \to s') = \min\left\{1, \frac{\pi(s') j_{-m}(s') g_{-m}(u')}{\pi(s) j_m(s) g_m(u)} \left|\frac{\partial(s', u')}{\partial(s, u)}\right|\right\}$$

- $j_m(s)$ Probability of choosing move m when in s

Moves



Toy Example 2

- Suppose we have two clipped Gaussian distributions
  - One Gaussian lives in $\mathbb{R}$, the other in $\mathbb{R}^2$
- We pretend as if we can not compute the area explicitly
- We wish to construct a RJMCMC sampler such that the chain spends more time in the domain where the area is higher

Toy Example 2

- If $s \in \mathbb{R}$

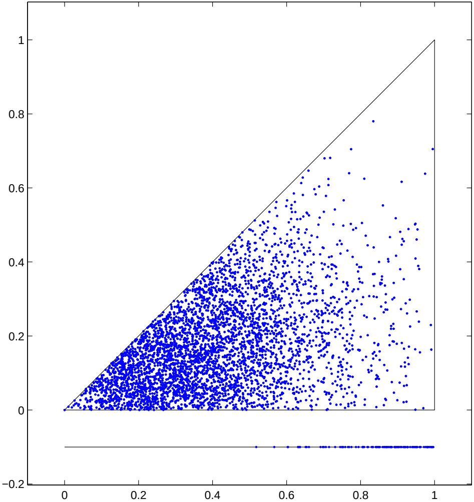
$$s \sim \frac{1}{Z_1}[0 < s < 1]\mathcal{N}(s; 1, \Sigma_1)$$

- If $s = (s_1, s_2) \in \mathbb{R}^2$

$$s \sim \frac{1}{Z_2}[0 < s_2 < s_1 < 1]\mathcal{N}(s; 0, \Sigma_2)$$

- Target distribution

$$\begin{aligned} p_1 &= Z_1/(Z_1 + Z_2) \qquad p_2 = Z_2/(Z_1 + Z_2) \\ \pi(s) &= \frac{1}{Z_1 + Z_2}[0 < s < 1]\mathcal{N}(s; 1, \Sigma_1)|_{s \in \mathbb{R}} \\ &\quad + \frac{1}{Z_1 + Z_2}[0 < s_2 < s_1 < 1]\mathcal{N}(s; 0, \Sigma_2)|_{s \in \mathbb{R}^2} \end{aligned}$$


1
0.8
0.6
0.4
0.2
0
−0.2
0
0.2
0.4
0.6
0.8
1

Toy Example 2

$1 \leftrightarrow 1$ In $\mathbb{R}$

$$s' \sim \mathcal{N}(s'; s, Q_1) = q(s'|s, m = (1 \leftrightarrow 1))$$

$2 \leftrightarrow 2$ In $\mathbb{R}^2$

$$(s'_1, s'_2) \sim \mathcal{N}((s'_1, s'_2); (s_1, s_2), Q_2)$$

$1 \leftrightarrow 2$ Transdimensional moves:

$$\begin{array}{ll} \mathbb{R} \to \mathbb{R}^2 & (s'_1, s'_2)^\top \sim \mathcal{N}(s'; \begin{pmatrix} s & 0 \end{pmatrix}^\top, Q_2) \\ \mathbb{R}^2 \to \mathbb{R} & s' \sim \mathcal{N}(s'; s_1, Q_1) \end{array}$$

Toy Example 2, Transdimensional move

- $\mathbb{R} \rightarrow \mathbb{R}^2$

$$
\begin{aligned}
s'_1 &= s + u_1 \\
s'_2 &= u_2
\end{aligned}
$$

- $\mathbb{R}^2 \rightarrow \mathbb{R}$, the reverse move

$$
s = s'_1 + u'
$$

- The transformation is

$$
\begin{aligned}
s'_1 &= s + u_1 \\
s'_2 &= u_2 \\
u' &= s - s'_1 = u_1
\end{aligned}
$$

Toy Example 2, Transdimensional move

- The Jacobian

$$
\begin{aligned}
\left|\frac{\partial(s_1', s_2', u')}{\partial(s, u_1, u_2)}\right| &= \left|\begin{array}{ccc} \partial s_1'/\partial s & \partial s_2'/\partial s & \partial u'/\partial s \\ \partial s_1'/\partial u_1 & \partial s_2'/\partial u_1 & \partial u'/\partial u_1 \\ \partial s_1'/\partial u_2 & \partial s_2'/\partial u_2 & \partial u'/\partial u_2 \end{array}\right| \\
&= \left|\begin{array}{ccc} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{array}\right| = |-1| = 1
\end{aligned}
$$

Toy Example 2, Transdimensional move

- Proposals

$$
\begin{aligned}
g_{1\to 2}(u_1, u_2) &= \mathcal{N}((u_1\ u_2)^\top; 0, Q_2) \\
g_{2\to 1}(u) &= \mathcal{N}(u; 0, Q_1)
\end{aligned}
$$

- The Jacobian

$$
\left|\frac{\partial(s', u')}{\partial(s, u)}\right| = \left|\partial s_1' s_1\right|
$$

Assignment

- Design the RJ MCMCM and Implement it for Toy Problem 2
- Calculate Z_1 and Z_2 numerically via Importance sampling and test the results of your RJ-MCMC implementation with the true values.
- (Optional) Design a RJMCMC algorithm for the sum of unknown number of dice example