Metropolis-Hastings for Mixtures of Conditional Distributions

Oleg Kiselyov
Tohoku University
oleg@okmij.org

```
p1c = do x ← dist bern 0.5
         y ← if x then condition 10 norm 10 1
                  else condition 10 norm 11 1
         return x
```

- What is the distribution of x?
- Can you write an MCMC sampling procedure for that distribution? What is the acceptance ratio formula?
- Why is y unused? Is conditioning a side-effect?

1 Why embedded conditioning matters

1.1 Computing the likelihood for the conditioned variable

Joint distribution on (x,y)

```
p1 = do x ← dist bern 0.5
        y ← if x then dist norm 10 1 else dist norm 11 1
        return (x,y)
```

Conditioning on y being 10

```
p1c' = do (x,y) ← p1
          observe (y==10)  -- Not a valid Hakaru10 statement!
          return x
```

Valid conditioning on y being 10

```
p1c'' = do (x,y) ← p1
           observe lh y 10  -- where to get y's likelihood from?
           return x
```

A better idea is to *push* the conditioning statement into p1, where y is computed and where its distribution is obvious. Hence we get p1c.

1.2 Modularity

If we are allowed to compose previously written models into new ones, what used to be 'top level' conditioning quickly ends up buried inside.

How to MCMC sample from models with embedded conditioning?

2 Warm-up: Models with branching

Deriving the correct Wingate et al. formula

```
p2 = do x ← dist bern a
        y ← if x then do {yt ← et; return yt} else do {yf ← ef; return yf}
        return (x,y)
```

Acceptance ratio for $s_1 \to s_2$

$$\alpha(s_1, s_2) = \min(1, r(s_2, s_1)) \qquad r(s_2, s_1) = \frac{\pi(s_2) q(s_2, s_1)}{\pi(s_1) q(s_1, s_2)}$$

($q(s_1, s_2)$: proposal kernel; $\pi(s)$ target density)

2.1 Wingate-like method

- A model (= program): a DAG of elementary random primitives (ERP) like bern in p2
- Each ERP is uniquely named (contra-Wingate)
- $|p|$: the number of *active* ERPs in p
- A sample from a program (= trace): a set of samples of all ERPs
- To sample from a program, we build a Markov Chain over the space of traces, by proposing an update to one ERP sample

2.2 MH sampling from p2

Current state	s_1: $(x{=}\text{true}, y{=}yt)$	$\pi(s_1) = \pi(x{=}\text{true})\ \delta(y{=}yt)\ \pi(et{=}yt)\ \pi(ef{=}yf)$
Proposed state	s_2: $(x{=}\text{false}, y{=}yf)$	$\pi(s_2)$ similarly

Proposal kernel $q(s_1, s_2)$

- x from all other eligible ERPs: $1/(1 + |et|)$.
- chose to update x from true to false: b_{tf}
- switch from $y{=}yt$ to $y{=}yf$: 1

$$r(s_2, s_1) = (1 - a)/a \cdot b_{ft}/b_{tf} \cdot (1 + |et|)/(1 + |ef|)$$

which is the correct (in the 3d revision) Wingate formula

3 Models with conditioning and branching

```
p3c = do x ← dist bern a
         y ← if x then do {yt ←et ; return yt}  -- now et, ef may have
                  else do {yf ←ef ; return yf}  -- conditioning!
         return (x,y)
```

To be explicit

```
p3 = do x ← dist bern a
        (y,z) ← if x then do {(yt,zt) ← et; return (yt,zt)}
                     else do {(yf,zf) ← ef; return (yf,zf)}
        return (x,y,z)
p3c = p3 conditioned on z=0
```

Consider the same transition as before s_1: $(x{=}\text{true}, y{=}yt) \rightsquigarrow s_2$: $(x{=}\text{false}, y{=}yf)$

$$\pi(s_1) = \pi(x{=}\text{true})\ \delta(y{=}yt)\ \pi(et{=}yt)\ \pi(ef{=}yf)\ \pi(zt{=}0)$$

Additional factor $\pi(zt{=}0)$!

Acceptance ratio

$$r(s_2, s_1) = \frac{\pi(x{=}\text{false})}{\pi(x{=}\text{true})} \frac{\pi(zf{=}0)}{\pi(zt{=}0)} \frac{b_{ft}}{b_{tf}} \frac{1+|et|}{1+|ef|}$$

Additional factor scoring the observation $z{=}0$ within the distributions of et and ef.

Submodels in conditional branches affect the acceptance ratio for the move to switch branches: conditioning is a side-effect!

4 Implementation

Hakaru10 http://okmij.org/ftp/kakuritu/Hakaru10/