CSci 8980: Advanced Topics in Graphical Models

Expectation Propagation

Instructor: Arindam Banerjee

October 26, 2007

Posterior Estimation

- Consider a Bayesian model

Posterior Estimation

- Consider a Bayesian model
  - Latent variable $\mathbf{u}$ with prior $P^{(0)}(\mathbf{u})$

Posterior Estimation

- Consider a Bayesian model
  - Latent variable $\mathbf{u}$ with prior $P^{(0)}(\mathbf{u})$
  - Observable D, such as $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$

Posterior Estimation

- Consider a Bayesian model
  - Latent variable $\mathbf{u}$ with prior $P^{(0)}(\mathbf{u})$
  - Observable D, such as $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$
- Quantities of interest

Posterior Estimation

- Consider a Bayesian model
  - Latent variable $\mathbf{u}$ with prior $P^{(0)}(\mathbf{u})$
  - Observable D, such as $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$
- Quantities of interest
  - Posterior over latent variable $P^{(0)}(\mathbf{u}|D)$

Posterior Estimation

- Consider a Bayesian model
  - Latent variable $\mathbf{u}$ with prior $P^{(0)}(\mathbf{u})$
  - Observable D, such as $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$
- Quantities of interest
  - Posterior over latent variable $P^{(0)}(\mathbf{u}|D)$
  - Likelihood of observation $P(D)$

Posterior Estimation

- Consider a Bayesian model
  - Latent variable $\mathbf{u}$ with prior $P^{(0)}(\mathbf{u})$
  - Observable D, such as $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$
- Quantities of interest
  - Posterior over latent variable $P^{(0)}(\mathbf{u}|D)$
  - Likelihood of observation $P(D)$
- For conjugate priors, posterior is in the same family

Posterior Estimation

- Consider a Bayesian model
  - Latent variable $\mathbf{u}$ with prior $P^{(0)}(\mathbf{u})$
  - Observable D, such as $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$
- Quantities of interest
  - Posterior over latent variable $P^{(0)}(\mathbf{u}|D)$
  - Likelihood of observation $P(D)$
- For conjugate priors, posterior is in the same family
- In general, it can be intractable

Posterior Estimation

- Consider a Bayesian model
  - Latent variable $\mathbf{u}$ with prior $P^{(0)}(\mathbf{u})$
  - Observable D, such as $\{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$
- Quantities of interest
  - Posterior over latent variable $P^{(0)}(\mathbf{u}|D)$
  - Likelihood of observation $P(D)$
- For conjugate priors, posterior is in the same family
- In general, it can be intractable
- What is the best approximation in the (prior) family?

Posterior Estimation (Contd.)

- The likelihood function often factorizes

$$P(D|\mathbf{u}) = \prod_{i=1}^{n} t_i(\mathbf{u})$$

Posterior Estimation (Contd.)

- The likelihood function often factorizes

$$P(D|\mathbf{u}) = \prod_{i=1}^{n} t_i(\mathbf{u})$$

- The true posterior may be intractable

$$P(\mathbf{u}|D) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} t_i(\mathbf{u})$$

Posterior Estimation (Contd.)

- The likelihood function often factorizes

$$P(D|\mathbf{u}) = \prod_{i=1}^{n} t_i(\mathbf{u})$$

- The true posterior may be intractable

$$P(\mathbf{u}|D) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} t_i(\mathbf{u})$$

- The normalizer Z is the same as the data likelihood, i.e.,

$$Z = \int_{\mathbf{u}} P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} t_i(\mathbf{u}) d\mathbf{u} = \int_{\mathbf{u}} P(D|\mathbf{u}) P^{(0)}(\mathbf{u}) d\mathbf{u} = P(D)$$

Posterior Estimation (Contd.)

- The likelihood function often factorizes

$$P(D|\mathbf{u}) = \prod_{i=1}^{n} t_i(\mathbf{u})$$

- The true posterior may be intractable

$$P(\mathbf{u}|D) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} t_i(\mathbf{u})$$

- The normalizer Z is the same as the data likelihood, i.e.,

$$Z = \int_{\mathbf{u}} P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} t_i(\mathbf{u}) d\mathbf{u} = \int_{\mathbf{u}} P(D|\mathbf{u}) P^{(0)}(\mathbf{u}) d\mathbf{u} = P(D)$$

- The two problems are closely related

Approximating the Posterior

- Assume prior $P^{(0)}(\mathbf{u})$ belongs to exponential family $\mathcal{F}$

$$P^{(0)}(\mathbf{u}) = \exp(\langle \theta_0, s(\mathbf{u}) \rangle - \psi(\theta))$$

Approximating the Posterior

- Assume prior $P^{(0)}(\mathbf{u})$ belongs to exponential family $\mathcal{F}$

$$P^{(0)}(\mathbf{u}) = \exp(\langle \theta_0, s(\mathbf{u}) \rangle - \psi(\theta))$$

- Let $Q(\mathbf{u}) \in \mathcal{F}$ be the best approximation to $P(\mathbf{u}|D)$

Approximating the Posterior

- Assume prior $P^{(0)}(\mathbf{u})$ belongs to exponential family $\mathcal{F}$

$$P^{(0)}(\mathbf{u}) = \exp(\langle \theta_0, s(\mathbf{u}) \rangle - \psi(\theta))$$

- Let $Q(\mathbf{u}) \in \mathcal{F}$ be the best approximation to $P(\mathbf{u}|D)$
- Tractably compute $Q(\mathbf{u})$ when $P(\mathbf{u}|D)$ is hard to compute

Approximating the Posterior

- Assume prior $P^{(0)}(\mathbf{u})$ belongs to exponential family $\mathcal{F}$

$$P^{(0)}(\mathbf{u}) = \exp(\langle \theta_0, s(\mathbf{u}) \rangle - \psi(\theta))$$

- Let $Q(\mathbf{u}) \in \mathcal{F}$ be the best approximation to $P(\mathbf{u}|D)$
- Tractably compute $Q(\mathbf{u})$ when $P(\mathbf{u}|D)$ is hard to compute
  - Approach 1: Assumed density filtering, online Bayesian learning

Approximating the Posterior

- Assume prior $P^{(0)}(\mathbf{u})$ belongs to exponential family $\mathcal{F}$

$$P^{(0)}(\mathbf{u}) = \exp(\langle \theta_0, s(\mathbf{u}) \rangle - \psi(\theta))$$

- Let $Q(\mathbf{u}) \in \mathcal{F}$ be the best approximation to $P(\mathbf{u}|D)$
- Tractably compute $Q(\mathbf{u})$ when $P(\mathbf{u}|D)$ is hard to compute
  - Approach 1: Assumed density filtering, online Bayesian learning
  - Approach 2: Expectation propagation

Assumed Density Filtering

- Start with an initial guess $Q(\mathbf{u}) = P^{(0)}(\mathbf{u})$

Assumed Density Filtering

- Start with an initial guess $Q(\mathbf{u}) = P^{(0)}(\mathbf{u})$
- Recall that

$$P(\mathbf{u}|D) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} t_i(\mathbf{u})$$

Assumed Density Filtering

- Start with an initial guess $Q(\mathbf{u}) = P^{(0)}(\mathbf{u})$
- Recall that

$$P(\mathbf{u}|D) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} t_i(\mathbf{u})$$

- At each step, update Q to incorporate one $t_i(\mathbf{u})$

Assumed Density Filtering

- Start with an initial guess $Q(\mathbf{u}) = P^{(0)}(\mathbf{u})$
- Recall that

$$P(\mathbf{u}|D) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} t_i(\mathbf{u})$$

- At each step, update Q to incorporate one $t_i(\mathbf{u})$
  - Compute the true Bayesian update

$$\hat{P}(\mathbf{u}) = \frac{t_i(\mathbf{u})Q(\mathbf{u})}{\int_{\mathbf{z}} t_i(\mathbf{z})Q(\mathbf{z})d\mathbf{z}}$$

Assumed Density Filtering

- Start with an initial guess $Q(\mathbf{u}) = P^{(0)}(\mathbf{u})$
- Recall that

$$P(\mathbf{u}|D) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} t_i(\mathbf{u})$$

- At each step, update Q to incorporate one $t_i(\mathbf{u})$
  - Compute the true Bayesian update

$$\hat{P}(\mathbf{u}) = \frac{t_i(\mathbf{u})Q(\mathbf{u})}{\int_{\mathbf{z}} t_i(\mathbf{z})Q(\mathbf{z})d\mathbf{z}}$$

  - Find $Q^{new} \in \mathcal{F}$ such that

$$Q^{new}(\mathbf{u}) = \underset{\tilde{Q} \in \mathcal{F}}{\operatorname{argmin}} \; KL(\hat{P}(u) \| \tilde{Q}(u))$$

Assumed Density Filtering

- Start with an initial guess $Q(\mathbf{u}) = P^{(0)}(\mathbf{u})$
- Recall that

$$P(\mathbf{u}|D) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} t_i(\mathbf{u})$$

- At each step, update Q to incorporate one $t_i(\mathbf{u})$
  - Compute the true Bayesian update

$$\hat{P}(\mathbf{u}) = \frac{t_i(\mathbf{u})Q(\mathbf{u})}{\int_{\mathbf{z}} t_i(\mathbf{z})Q(\mathbf{z})d\mathbf{z}}$$

  - Find $Q^{new} \in \mathcal{F}$ such that

$$Q^{new}(\mathbf{u}) = \operatorname*{argmin}_{\tilde{Q} \in \mathcal{F}} KL(\hat{P}(u) \| \tilde{Q}(u))$$

- Maximum likelihood estimate with $\hat{P}$ as the true distribution

Assumed Density Filtering (Contd.)

- To obtain Q^{new} it is sufficient to do moment matching

$$\mu_{new} = E_{\hat{P}}[s(\mathbf{u})]$$

Assumed Density Filtering (Contd.)

- To obtain Q^{new} it is sufficient to do moment matching

$$\mu_{new} = E_{\hat{P}}[s(\mathbf{u})]$$

- For each factor $t_i(\mathbf{u})$

Assumed Density Filtering (Contd.)

- To obtain Q^{new} it is sufficient to do moment matching

$$\mu_{new} = E_{\hat{P}}[s(\mathbf{u})]$$

- For each factor $t_i(\mathbf{u})$
  - Compute the means (moments) of $\hat{P}(\mathbf{u}) \propto t_i(\mathbf{u})Q(\mathbf{u})$

Assumed Density Filtering (Contd.)

- To obtain Q^{new} it is sufficient to do moment matching

$$\mu_{new} = E_{\hat{P}}[s(\mathbf{u})]$$

- For each factor $t_i(\mathbf{u})$
  - Compute the means (moments) of $\hat{P}(\mathbf{u}) \propto t_i(\mathbf{u})Q(\mathbf{u})$
  - Pick $Q^{new} \in \mathcal{F}$ with these mean parameters

ADF: An Alternative Viewpoint

- For a single factor $t_i(\mathbf{u})$

ADF: An Alternative Viewpoint

- For a single factor $t_i(\mathbf{u})$
  - The true posterior $\hat{P}(\mathbf{u}) \propto t_i(\mathbf{u})Q(\mathbf{u})$

ADF: An Alternative Viewpoint

- For a single factor $t_i(\mathbf{u})$
  - The true posterior $\hat{P}(\mathbf{u}) \propto t_i(\mathbf{u})Q(\mathbf{u})$
  - The approximate posterior $Q^{new}(\mathbf{u}) \propto \tilde{t}_i(\mathbf{u})Q(\mathbf{u})$

ADF: An Alternative Viewpoint

- For a single factor $t_i(\mathbf{u})$
  - The true posterior $\hat{P}(\mathbf{u}) \propto t_i(\mathbf{u})Q(\mathbf{u})$
  - The approximate posterior $Q^{new}(\mathbf{u}) \propto \tilde{t}_i(\mathbf{u})Q(\mathbf{u})$
  - The factor $\tilde{t}_i(\mathbf{u}) \propto Q^{new}(\mathbf{u})/Q(\mathbf{u})$

ADF: An Alternative Viewpoint

- For a single factor $t_i(\mathbf{u})$
  - The true posterior $\hat{P}(\mathbf{u}) \propto t_i(\mathbf{u})Q(\mathbf{u})$
  - The approximate posterior $Q^{new}(\mathbf{u}) \propto \tilde{t}_i(\mathbf{u})Q(\mathbf{u})$
  - The factor $\tilde{t}_i(\mathbf{u}) \propto Q^{new}(\mathbf{u})/Q(\mathbf{u})$
- In general, after a pass through all factors

$$Q(\mathbf{u}) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})$$

ADF: An Alternative Viewpoint

- For a single factor $t_i(\mathbf{u})$
  - The true posterior $\hat{P}(\mathbf{u}) \propto t_i(\mathbf{u})Q(\mathbf{u})$
  - The approximate posterior $Q^{new}(\mathbf{u}) \propto \tilde{t}_i(\mathbf{u})Q(\mathbf{u})$
  - The factor $\tilde{t}_i(\mathbf{u}) \propto Q^{new}(\mathbf{u})/Q(\mathbf{u})$
- In general, after a pass through all factors

$$Q(\mathbf{u}) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})$$

- Algo: Set $\tilde{t}_i = 1, \forall i$. For each factor $t_i(\mathbf{u})$

ADF: An Alternative Viewpoint

- For a single factor $t_i(\mathbf{u})$
  - The true posterior $\hat{P}(\mathbf{u}) \propto t_i(\mathbf{u})Q(\mathbf{u})$
  - The approximate posterior $Q^{new}(\mathbf{u}) \propto \tilde{t}_i(\mathbf{u})Q(\mathbf{u})$
  - The factor $\tilde{t}_i(\mathbf{u}) \propto Q^{new}(\mathbf{u})/Q(\mathbf{u})$
- In general, after a pass through all factors

$$Q(\mathbf{u}) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})$$

- Algo: Set $\tilde{t}_i = 1, \forall i$. For each factor $t_i(\mathbf{u})$
  - Compute Q^{new} with $\mu^{new} = E_{\hat{P}}[s(\mathbf{u})]$

ADF: An Alternative Viewpoint

- For a single factor $t_i(\mathbf{u})$
  - The true posterior $\hat{P}(\mathbf{u}) \propto t_i(\mathbf{u})Q(\mathbf{u})$
  - The approximate posterior $Q^{new}(\mathbf{u}) \propto \tilde{t}_i(\mathbf{u})Q(\mathbf{u})$
  - The factor $\tilde{t}_i(\mathbf{u}) \propto Q^{new}(\mathbf{u})/Q(\mathbf{u})$
- In general, after a pass through all factors

$$Q(\mathbf{u}) \propto P^{(0)}(\mathbf{u}) \prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})$$

- Algo: Set $\tilde{t}_i = 1, \forall i$. For each factor $t_i(\mathbf{u})$
  - Compute Q^{new} with $\mu^{new} = E_{\hat{P}}[s(\mathbf{u})]$
  - Set $\tilde{t}_i^{new}(\mathbf{u}) \propto Q^{new}(\mathbf{u})/Q(\mathbf{u})$

Issues with ADF

- ADF makes one pass through the data

Issues with ADF

- ADF makes one pass through the data
  - Q is updated once for each factor

Issues with ADF

- ADF makes one pass through the data
  - Q is updated once for each factor
  - Equivalently, $\tilde{t}_i(\mathbf{u})$ is updated once

Issues with ADF

- ADF makes one pass through the data
  - Q is updated once for each factor
  - Equivalently, $\tilde{t}_i(\mathbf{u})$ is updated once
- Issues with ADF

Issues with ADF

- ADF makes one pass through the data
    - Q is updated once for each factor
    - Equivalently, $\tilde{t}_i(\mathbf{u})$ is updated once
- Issues with ADF
    - Earlier approximations $\tilde{t}_i(\mathbf{u})$ may be poor

Issues with ADF

- ADF makes one pass through the data
  - Q is updated once for each factor
  - Equivalently, $\tilde{t}_i(\mathbf{u})$ is updated once
- Issues with ADF
  - Earlier approximations $\tilde{t}_i(\mathbf{u})$ may be poor
  - No way of going back and fixing the earlier approximations

Issues with ADF

- ADF makes one pass through the data
  - Q is updated once for each factor
  - Equivalently, $\tilde{t}_i(\mathbf{u})$ is updated once
- Issues with ADF
  - Earlier approximations $\tilde{t}_i(\mathbf{u})$ may be poor
  - No way of going back and fixing the earlier approximations
  - Depends on the order in which data is processed

Issues with ADF

- ADF makes one pass through the data
  - Q is updated once for each factor
  - Equivalently, $\tilde{t}_i(\mathbf{u})$ is updated once
- Issues with ADF
  - Earlier approximations $\tilde{t}_i(\mathbf{u})$ may be poor
  - No way of going back and fixing the earlier approximations
  - Depends on the order in which data is processed
- In principle, $\tilde{t}_i$ can be updated multiple times

Issues with ADF

- ADF makes one pass through the data
  - Q is updated once for each factor
  - Equivalently, $\tilde{t}_i(\mathbf{u})$ is updated once
- Issues with ADF
  - Earlier approximations $\tilde{t}_i(\mathbf{u})$ may be poor
  - No way of going back and fixing the earlier approximations
  - Depends on the order in which data is processed
- In principle, $\tilde{t}_i$ can be updated multiple times
- EP effectively extends ADF allowing multiple passes

Expectation Propagation

- Initialize the term approximations $\tilde{t}_i(\mathbf{u})$

Expectation Propagation

- Initialize the term approximations $\tilde{t}_i(\mathbf{u})$
- Compute the posterior

$$Q(\mathbf{u}) = \frac{\prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})}{\int_{\mathbf{z}} \prod_{i=1}^{n} \tilde{t}_i(\mathbf{z}) d\mathbf{z}}$$

Expectation Propagation

- Initialize the term approximations $\tilde{t}_i(\mathbf{u})$
- Compute the posterior

$$Q(\mathbf{u}) = \frac{\prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})}{\int_{\mathbf{z}} \prod_{i=1}^{n} \tilde{t}_i(\mathbf{z}) d\mathbf{z}}$$

- Until all $\tilde{t}_i$ converge

Expectation Propagation

- Initialize the term approximations $\tilde{t}_i(\mathbf{u})$
- Compute the posterior

$$Q(\mathbf{u}) = \frac{\prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})}{\int_{\mathbf{z}} \prod_{i=1}^{n} \tilde{t}_i(\mathbf{z}) d\mathbf{z}}$$

- Until all $\tilde{t}_i$ converge
  - Choose a $\tilde{t}_i(\mathbf{u})$ to refine

Expectation Propagation

- Initialize the term approximations $\tilde{t}_i(\mathbf{u})$
- Compute the posterior

$$Q(\mathbf{u}) = \frac{\prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})}{\int_{\mathbf{z}} \prod_{i=1}^{n} \tilde{t}_i(\mathbf{z}) d\mathbf{z}}$$

- Until all $\tilde{t}_i$ converge
  - Choose a $\tilde{t}_i(\mathbf{u})$ to refine
  - Remove $\tilde{t}_i(\mathbf{u})$ from $Q(\mathbf{u})$ to get 'old' posterior

$$Q^{\setminus i}(\mathbf{u}) \propto Q(\mathbf{u}) / \tilde{t}_i(\mathbf{u})$$

Expectation Propagation

- Initialize the term approximations $\tilde{t}_i(\mathbf{u})$
- Compute the posterior

$$Q(\mathbf{u}) = \frac{\prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})}{\int_{\mathbf{z}} \prod_{i=1}^{n} \tilde{t}_i(\mathbf{z}) d\mathbf{z}}$$

- Until all $\tilde{t}_i$ converge
  - Choose a $\tilde{t}_i(\mathbf{u})$ to refine
  - Remove $\tilde{t}_i(\mathbf{u})$ from $Q(\mathbf{u})$ to get 'old' posterior

$$Q^{\setminus i}(\mathbf{u}) \propto Q(\mathbf{u}) / \tilde{t}_i(\mathbf{u})$$

  - Construct $\hat{P} \propto t_i(\mathbf{u}) Q^{\setminus i}(\mathbf{u})$

Expectation Propagation

- Initialize the term approximations $\tilde{t}_i(\mathbf{u})$
- Compute the posterior

$$Q(\mathbf{u}) = \frac{\prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})}{\int_{\mathbf{z}} \prod_{i=1}^{n} \tilde{t}_i(\mathbf{z}) d\mathbf{z}}$$

- Until all $\tilde{t}_i$ converge
  - Choose a $\tilde{t}_i(\mathbf{u})$ to refine
  - Remove $\tilde{t}_i(\mathbf{u})$ from $Q(\mathbf{u})$ to get 'old' posterior

$$Q^{\,i}(\mathbf{u}) \propto Q(\mathbf{u})/\tilde{t}_i(\mathbf{u})$$

  - Construct $\hat{P} \propto t_i(\mathbf{u}) Q^{\,i}(\mathbf{u})$
  - Get Q^{new} with $\mu^{new} = E_{\hat{P}}[s(\mathbf{u})]$ and normalizer Z_i

Expectation Propagation

- Initialize the term approximations $\tilde{t}_i(\mathbf{u})$
- Compute the posterior

$$Q(\mathbf{u}) = \frac{\prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})}{\int_{\mathbf{z}} \prod_{i=1}^{n} \tilde{t}_i(\mathbf{z}) d\mathbf{z}}$$

- Until all $\tilde{t}_i$ converge
    - Choose a $\tilde{t}_i(\mathbf{u})$ to refine
    - Remove $\tilde{t}_i(\mathbf{u})$ from $Q(\mathbf{u})$ to get 'old' posterior

$$Q^{\setminus i}(\mathbf{u}) \propto Q(\mathbf{u}) / \tilde{t}_i(\mathbf{u})$$

    - Construct $\hat{P} \propto t_i(\mathbf{u}) Q^{\setminus i}(\mathbf{u})$
    - Get Q^{new} with $\mu^{new} = E_{\hat{P}}[s(\mathbf{u})]$ and normalizer Z_i
    - Update $\tilde{t}_i(\mathbf{u}) = Z_i Q^{new}(\mathbf{u}) / Q^{\setminus i}(\mathbf{u})$

Expectation Propagation

- Initialize the term approximations $\tilde{t}_i(\mathbf{u})$
- Compute the posterior

$$Q(\mathbf{u}) = \frac{\prod_{i=1}^{n} \tilde{t}_i(\mathbf{u})}{\int_{\mathbf{z}} \prod_{i=1}^{n} \tilde{t}_i(\mathbf{z}) d\mathbf{z}}$$

- Until all $\tilde{t}_i$ converge
  - Choose a $\tilde{t}_i(\mathbf{u})$ to refine
  - Remove $\tilde{t}_i(\mathbf{u})$ from $Q(\mathbf{u})$ to get 'old' posterior

$$Q^{\setminus i}(\mathbf{u}) \propto Q(\mathbf{u}) / \tilde{t}_i(\mathbf{u})$$

  - Construct $\hat{P} \propto t_i(\mathbf{u}) Q^{\setminus i}(\mathbf{u})$
  - Get Q^{new} with $\mu^{new} = E_{\hat{P}}[s(\mathbf{u})]$ and normalizer Z_i
  - Update $\tilde{t}_i(\mathbf{u}) = Z_i Q^{new}(\mathbf{u}) / Q^{\setminus i}(\mathbf{u})$
- Estimate the data likelihood as

$$P(D) \approx \int_{\mathbf{z}} \prod_{i=1}^{n} \tilde{t}_i(\mathbf{z}) d\mathbf{z}$$

Experiments

- The clutter problem:

$$
\begin{aligned}
p(\mathbf{u}) &= \mathcal{N}(\mathbf{u}; \mathbf{0}, \mathbf{100}\mathbb{I}) \\
p(\mathbf{x}|\mathbf{u}) &= (1-w)\mathcal{N}(\mathbf{x}; \mathbf{u}, \mathbb{I}) + w\mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{100}\mathbb{I})
\end{aligned}
$$

Experiments

- The clutter problem:

$$
\begin{aligned}
p(\mathbf{u}) &= \mathcal{N}(\mathbf{u}; \mathbf{0}, \mathbf{100}\mathbb{I}) \\
p(\mathbf{x}|\mathbf{u}) &= (1-w)\mathcal{N}(\mathbf{x}; \mathbf{u}, \mathbb{I}) + w\mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{100}\mathbb{I})
\end{aligned}
$$

- For a set of observations $D = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$

$$
p(\mathbf{u}, D) = p(\mathbf{u}) \prod_{j=1}^{n} p(\mathbf{x}_j|\mathbf{u})
$$

Experiments

- The clutter problem:

$$
\begin{aligned}
p(\mathbf{u}) &= \mathcal{N}(\mathbf{u}; \mathbf{0}, \mathbf{100}\mathbb{I}) \\
p(\mathbf{x}|\mathbf{u}) &= (1-w)\mathcal{N}(\mathbf{x}; \mathbf{u}, \mathbb{I}) + w\mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{100}\mathbb{I})
\end{aligned}
$$

- For a set of observations $D = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$

$$
p(\mathbf{u}, D) = p(\mathbf{u}) \prod_{j=1}^{n} p(\mathbf{x}_j|\mathbf{u})
$$

- Evaluation

Experiments

- The clutter problem:

$$
\begin{aligned}
p(\mathbf{u}) &= \mathcal{N}(\mathbf{u}; \mathbf{0}, \mathbf{100}\mathbb{I}) \\
p(\mathbf{x}|\mathbf{u}) &= (1-w)\mathcal{N}(\mathbf{x}; \mathbf{u}, \mathbb{I}) + w\mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{100}\mathbb{I})
\end{aligned}
$$

- For a set of observations $D = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$

$$
p(\mathbf{u}, D) = p(\mathbf{u}) \prod_{j=1}^{n} p(\mathbf{x}_j|\mathbf{u})
$$

- Evaluation
  - Evidence/likelihood $p(D) = \int_{\mathbf{u}} p(\mathbf{u}, D) d\mathbf{u}$

Experiments

- The clutter problem:

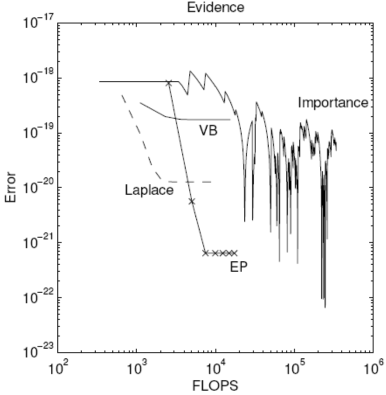
$$p(\mathbf{u}) = \mathcal{N}(\mathbf{u}; \mathbf{0}, \mathbf{100}\mathbb{I})$$
$$p(\mathbf{x}|\mathbf{u}) = (1-w)\mathcal{N}(\mathbf{x}; \mathbf{u}, \mathbb{I}) + w\mathcal{N}(\mathbf{x}; \mathbf{0}, \mathbf{100}\mathbb{I})$$

- For a set of observations $D = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$

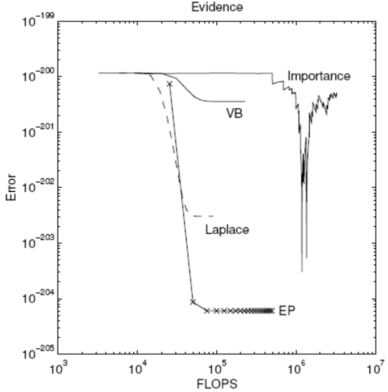
$$p(\mathbf{u}, D) = p(\mathbf{u}) \prod_{j=1}^{n} p(\mathbf{x}_j|\mathbf{u})$$

- Evaluation
  - Evidence/likelihood $p(D) = \int_{\mathbf{u}} p(\mathbf{u}, D) d\mathbf{u}$
  - Posterior mean $E[\mathbf{u}|D] = \int_{\mathbf{u}} \mathbf{u} p(\mathbf{u}|D) d\mathbf{u}$

Results: Likelihood $P(D)$

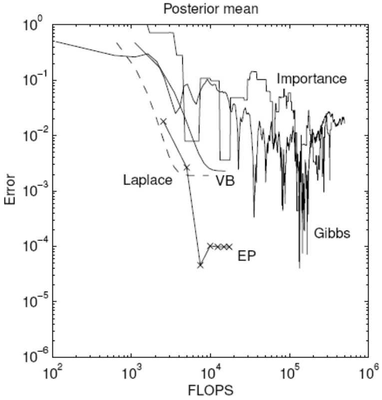


$n = 20$

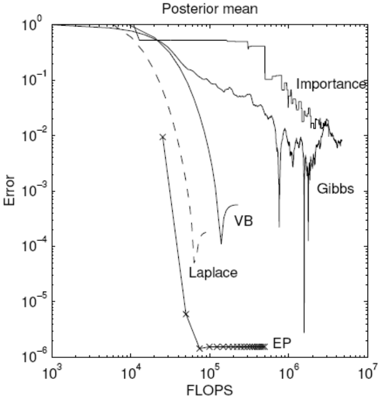


$n = 200$

Results: Posterior Mean $E[\mathbf{u}|D]$



$n = 20$



$n = 200$

Results: Complex Posterior