Approximate Bayesian Inference for Geophysical Inverse Problems

Martin Jullum

University of Oslo

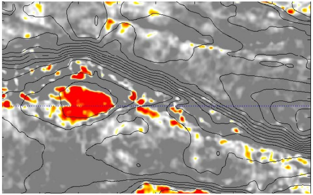
martinju@math.uio.no

May 25, 2013

Outline

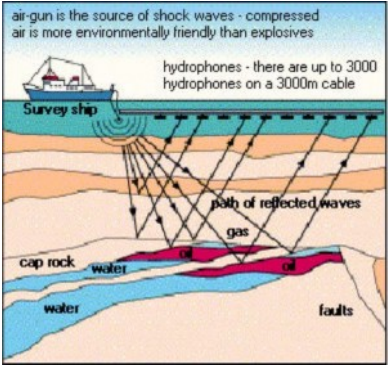
Oil and gas exploration

- Hard to determine where oil and gas are located
- Costs 200 - 800 million NOK to drill a exploration well
- Massive profit if one finds hydrocarbon reservoirs - large loss if not
  - Needs to estimate the probability of hydrocarbons



Seismic survey

- A source sends out shock waves
- Shock waves are reflected in the contrasts between the different facies (rock types)
- Receivers measure properties of smoothed reflected wave signals: Seismic data



Problem setup

- The behavior of the reflected waves depends on the contrast facies
- Seismic data can be processed to tell us about the material properties
- Different facies have different material properties

- We would like to estimate the facies probability distribution at various locations below the seabed

Relations:

- $Z = XY + \epsilon$, for some error term ϵ
  - Z: Seismic Data
  - X: Data link (design matrix/dynamics)
  - Y: Rock material characteristics
- θ: Facies categories

Problem setup

- The behavior of the reflected waves depends on the contrast facies
- Seismic data can be processed to tell us about the material properties
- Different facies have different material properties

- We would like to estimate the facies probability distribution at various locations below the seabed

Relations:

- $Z = XY + \epsilon$, for some error term ϵ
  - Z: Seismic Data
  - X: Data link (design matrix/dynamics)
  - Y: Rock material characteristics
- θ: Facies categories

Problem setup

- The behavior of the reflected waves depends on the contrast facies
- Seismic data can be processed to tell us about the material properties
- Different facies have different material properties

- We would like to estimate the facies probability distribution at various locations below the seabed

Relations:

- $Z = XY + \epsilon$, for some error term ϵ
  - Z: Seismic Data
  - X: Data link (design matrix/dynamics)
  - Y: Rock material characteristics
- θ: Facies categories

Hierarchical model setup

- $p(Z|Y)$: Seismic data model
- $p(Y|\theta)$: Latent rock material physics model
- $p(\theta)$: Categorical facies model

- The forward problem: Determine Z from θ
- The inverse problem we are interested in: Determine θ from Z: I.e. the standard Bayesian problem of finding the posterior $p(\theta|Z)$.

- **Main problem**
  - Great number of surface locations m, each with a large number of vertical positions - massive amounts of data
  - The wavelet smoothen the reflected signal data
  - $\dim(\theta)$ is enormous

Hierarchical model setup

- $p(Z|Y)$: Seismic data model
- $p(Y|\theta)$: Latent rock material physics model
- $p(\theta)$: Categorical facies model

- The forward problem: Determine Z from θ
- The inverse problem we are interested in: Determine θ from Z: I.e. the standard Bayesian problem of finding the posterior $p(\theta|Z)$.

- **Main problem**
  - Great number of surface locations m, each with a large number of vertical positions - massive amounts of data
  - The wavelet smoothen the reflected signal data
  - $\dim(\theta)$ is enormous

Hierarchical model setup

- $p(Z|Y)$: Seismic data model
- $p(Y|\theta)$: Latent rock material physics model
- $p(\theta)$: Categorical facies model

- The forward problem: Determine Z from θ
- The inverse problem we are interested in: Determine θ from Z: I.e. the standard Bayesian problem of finding the posterior $p(\theta|Z)$.

- **Main problem**
  - Great number of surface locations m, each with a large number of vertical positions - massive amounts of data
  - The wavelet smoothen the reflected signal data
  - $\dim(\theta)$ is enormous

How do we solve this?

- The simplest reasonable assumptions:
  - $p(Z|Y,\theta) = p(Z|Y)$ is Gaussian
  - $p(Y|\theta)$ is Gaussian and $p(Y_i|\theta) = p(Y_i|\theta_i)$ for some location i is also Gaussian
  - $p(\theta)$ is categorical

How to find/approximate $p(\theta|Z)$:

- Exact calculation - possible in theory, but not in practice
  - $p(Y)$ is a Gaussian mixture which returns an exact mathematical formulae for $p(\theta|Z)$
  - The calculation requires summation over θ - impossible due to large $\dim(\theta)$
- MCMC - similar problem as exact computation
- INLA - assumptions are not completely fulfilled
  - $p(Y|\theta)$ Gaussian: OK
  - $p(Z|Y,\theta) = \prod_{i=1}^{n} p(Z_i|\eta_i,\theta)$, for linear predictor η_i of the Y_is: Almost OK
  - $\dim(\theta)$ small: NO!

How do we solve this?

- The simplest reasonable assumptions:
  - $p(Z|Y,\theta) = p(Z|Y)$ is Gaussian
  - $p(Y|\theta)$ is Gaussian and $p(Y_i|\theta) = p(Y_i|\theta_i)$ for some location i is also Gaussian
  - $p(\theta)$ is categorical

How to find/approximate $p(\theta|Z)$:

- Exact calculation - possible in theory, but not in practice
  - $p(Y)$ is a Gaussian mixture which returns an exact mathematical formulae for $p(\theta|Z)$
  - The calculation requires summation over θ - impossible due to large $\dim(\theta)$
- MCMC - similar problem as exact computation
- INLA - assumptions are not completely fulfilled
  - $p(Y|\theta)$ Gaussian: OK
  - $p(Z|Y,\theta) = \prod_{i=1}^{n} p(Z_i|\eta_i,\theta)$, for linear predictor η_i of the Y_is: Almost OK
  - $\dim(\theta)$ small: NO!

How do we solve this?

- The simplest reasonable assumptions:
  - $p(Z|Y,\theta) = p(Z|Y)$ is Gaussian
  - $p(Y|\theta)$ is Gaussian and $p(Y_i|\theta) = p(Y_i|\theta_i)$ for some location i is also Gaussian
  - $p(\theta)$ is categorical

How to find/approximate $p(\theta|Z)$:

- Exact calculation - possible in theory, but not in practice
  - $p(Y)$ is a Gaussian mixture which returns an exact mathematical formulae for $p(\theta|Z)$
  - The calculation requires summation over θ - impossible due to large $\dim(\theta)$
- MCMC - similar problem as exact computation
- INLA - assumptions are not completely fulfilled
  - $p(Y|\theta)$ Gaussian: OK
  - $p(Z|Y,\theta) = \prod_{i=1}^{n} p(Z_i|\eta_i,\theta)$, for linear predictor η_i of the Y_is: Almost OK
  - $\dim(\theta)$ small: NO!

How do we solve this?

- The simplest reasonable assumptions:
  - $p(Z|Y,\theta) = p(Z|Y)$ is Gaussian
  - $p(Y|\theta)$ is Gaussian and $p(Y_i|\theta) = p(Y_i|\theta_i)$ for some location i is also Gaussian
  - $p(\theta)$ is categorical

How to find/approximate $p(\theta|Z)$:

- Exact calculation - possible in theory, but not in practice
  - $p(Y)$ is a Gaussian mixture which returns an exact mathematical formulae for $p(\theta|Z)$
  - The calculation requires summation over θ - impossible due to large $\dim(\theta)$
- MCMC - similar problem as exact computation
- INLA - assumptions are not completely fulfilled
  - $p(Y|\theta)$ Gaussian: OK
  - $p(Z|Y,\theta) = \prod_{i=1}^{n} p(Z_i|\eta_i,\theta)$, for linear predictor η_i of the Y_is: Almost OK
  - $\dim(\theta)$ small: NO!

How do we solve this?

- The simplest reasonable assumptions:
  - $p(Z|Y,\theta) = p(Z|Y)$ is Gaussian
  - $p(Y|\theta)$ is Gaussian and $p(Y_i|\theta) = p(Y_i|\theta_i)$ for some location i is also Gaussian
  - $p(\theta)$ is categorical

How to find/approximate $p(\theta|Z)$:

- Exact calculation - possible in theory, but not in practice
  - $p(Y)$ is a Gaussian mixture which returns an exact mathematical formulae for $p(\theta|Z)$
  - The calculation requires summation over θ - impossible due to large $\dim(\theta)$
- MCMC - similar problem as exact computation
- INLA - assumptions are not completely fulfilled
  - $p(Y|\theta)$ Gaussian: OK
  - $p(Z|Y,\theta) = \prod_{i=1}^{n} p(Z_i|\eta_i,\theta)$, for linear predictor η_i of the Y_is: Almost OK
  - $\dim(\theta)$ small: NO!

How do we solve this?

- The simplest reasonable assumptions:
  - $p(Z|Y,\theta) = p(Z|Y)$ is Gaussian
  - $p(Y|\theta)$ is Gaussian and $p(Y_i|\theta) = p(Y_i|\theta_i)$ for some location i is also Gaussian
  - $p(\theta)$ is categorical

How to find/approximate $p(\theta|Z)$:

- Exact calculation - possible in theory, but not in practice
  - $p(Y)$ is a Gaussian mixture which returns an exact mathematical formulae for $p(\theta|Z)$
  - The calculation requires summation over θ - impossible due to large $\dim(\theta)$
- MCMC - similar problem as exact computation
- INLA - assumptions are not completely fulfilled
  - $p(Y|\theta)$ Gaussian: OK
  - $p(Z|Y,\theta) = \prod_{i=1}^{n} p(Z_i|\eta_i,\theta)$, for linear predictor η_i of the Y_is: Almost OK
  - $\dim(\theta)$ small: NO!

How do we solve this?

- The simplest reasonable assumptions:
  - $p(Z|Y,\theta) = p(Z|Y)$ is Gaussian
  - $p(Y|\theta)$ is Gaussian and $p(Y_i|\theta) = p(Y_i|\theta_i)$ for some location i is also Gaussian
  - $p(\theta)$ is categorical

How to find/approximate $p(\theta|Z)$:

- Exact calculation - possible in theory, but not in practice
  - $p(Y)$ is a Gaussian mixture which returns an exact mathematical formulae for $p(\theta|Z)$
  - The calculation requires summation over θ - impossible due to large $\dim(\theta)$
- MCMC - similar problem as exact computation
- INLA - assumptions are not completely fulfilled
  - $p(Y|\theta)$ Gaussian: OK
  - $p(Z|Y,\theta) = \prod_{i=1}^{n} p(Z_i|\eta_i,\theta)$, for linear predictor η_i of the Y_is: Almost OK
  - $\dim(\theta)$ small: NO!

Locally focused approximation

Let us focus on $p(\theta_i|Z)$, that is the posterior in one location only.

- Let p_G denote distributions under the Gaussian assumption.
- If $p(Y) = p_G(Y)$, then $p(Y_i) = p_G(Y_i)$ and $p(Y_i|Z) = p_G(Y_i|Z)$ as well. Then

$$\begin{aligned} p(\theta_i|Z) &\propto \int p(Z, Y_i, \theta_i) dY_i = p(\theta_i) \int p(Z|Y_i) p(Y_i|\theta_i) dY_i \\ &\propto p(\theta_i) \int \frac{p_G(Y_i|Z)}{p_G(Y_i)} p(Y_i|\theta_i) dY_i \end{aligned}$$

- The integrand consist of only Gaussians, so this is easy.
- $p(Y) = p_G(Y)$ is convenient, but generally incorrect

Locally focused approximation

Let us focus on $p(\theta_i|Z)$, that is the posterior in one location only.

- Let p_G denote distributions under the Gaussian assumption.
- If $p(Y) = p_G(Y)$, then $p(Y_i) = p_G(Y_i)$ and $p(Y_i|Z) = p_G(Y_i|Z)$ as well. Then

$$p(\theta_i|Z) \propto \int p(Z, Y_i, \theta_i) dY_i = p(\theta_i) \int p(Z|Y_i) p(Y_i|\theta_i) dY_i$$
$$\propto p(\theta_i) \int \frac{p_G(Y_i|Z)}{p_G(Y_i)} p(Y_i|\theta_i) dY_i$$

- The integrand consist of only Gaussians, so this is easy.
- $p(Y) = p_G(Y)$ is convenient, but generally incorrect

Locally focused approximation

Let us focus on $p(\theta_i|Z)$, that is the posterior in one location only.

- Let p_G denote distributions under the Gaussian assumption.
- If $p(Y) = p_G(Y)$, then $p(Y_i) = p_G(Y_i)$ and $p(Y_i|Z) = p_G(Y_i|Z)$ as well. Then

$$p(\theta_i|Z) \propto \int p(Z, Y_i, \theta_i) dY_i = p(\theta_i) \int p(Z|Y_i) p(Y_i|\theta_i) dY_i$$
$$\propto p(\theta_i) \int \frac{p_G(Y_i|Z)}{p_G(Y_i)} p(Y_i|\theta_i) dY_i$$

- The integrand consist of only Gaussians, so this is easy.
- $p(Y) = p_G(Y)$ is convenient, but generally incorrect

Locally focused approximation

Let us focus on $p(\theta_i|Z)$, that is the posterior in one location only.

- Let p_G denote distributions under the Gaussian assumption.
- If $p(Y) = p_G(Y)$, then $p(Y_i) = p_G(Y_i)$ and $p(Y_i|Z) = p_G(Y_i|Z)$ as well. Then

$$
\begin{aligned}
p(\theta_i|Z) &\propto \int p(Z, Y_i, \theta_i) dY_i = p(\theta_i) \int p(Z|Y_i) p(Y_i|\theta_i) dY_i \\
&\propto p(\theta_i) \int \frac{p_G(Y_i|Z)}{p_G(Y_i)} p(Y_i|\theta_i) dY_i
\end{aligned}
$$

- The integrand consist of only Gaussians, so this is easy.
- $p(Y) = p_G(Y)$ is convenient, but generally incorrect

Locally focused approximation

Let us focus on $p(\theta_i|Z)$, that is the posterior in one location only.

- Let p_G denote distributions under the Gaussian assumption.
- If $p(Y) = p_G(Y)$, then $p(Y_i) = p_G(Y_i)$ and $p(Y_i|Z) = p_G(Y_i|Z)$ as well. Then

$$
\begin{aligned}
p(\theta_i|Z) &\propto \int p(Z, Y_i, \theta_i) dY_i = p(\theta_i) \int p(Z|Y_i) p(Y_i|\theta_i) dY_i \\
&\propto p(\theta_i) \int \frac{p_G(Y_i|Z)}{p_G(Y_i)} p(Y_i|\theta_i) dY_i
\end{aligned}
$$

- The integrand consist of only Gaussians, so this is easy.
- $p(Y) = p_G(Y)$ is convenient, but generally incorrect

Locally focused approximation from a different perspective

We have exactly that

$$p(Y|Z) \propto p_G(Y|Z)\frac{p(Y)}{p_G(Y)}.$$

Thus, if follows that

$$\begin{aligned} p(\theta_i|Z) &\propto p(\theta_i)\int p(Y|Z)p(Y)p(Y|\theta_i)dY \\ &\propto p(\theta_i)\int \frac{p_G(Y|Z)}{p_G(Y)}p(Y|\theta_i)dY \\ &= \cdots = p(\theta_i)\int \frac{p_G(Y_i|Z)}{p_G(Y_i)}p(Y_i|\theta)K(Y_i)dY_i \end{aligned}$$

where

$$K(Y_i) = \int \frac{p_G(Y_{-i}|Z, Y_i)}{p_G(Y_{-i}|Y_i)}p(Y_{-i}|\theta_i, Y_i)dY_{-i}$$

Locally focused approximation from a different perspective

We have exactly that

$$p(Y|Z) \propto p_G(Y|Z)\frac{p(Y)}{p_G(Y)}.$$

Thus, if follows that

$$\begin{aligned} p(\theta_i|Z) &\propto p(\theta_i)\int p(Y|Z)p(Y)p(Y|\theta_i)dY \\ &\propto p(\theta_i)\int \frac{p_G(Y|Z)}{p_G(Y)}p(Y|\theta_i)dY \\ &= \cdots = p(\theta_i)\int \frac{p_G(Y_i|Z)}{p_G(Y_i)}p(Y_i|\theta)K(Y_i)dY_i \end{aligned}$$

where

$$K(Y_i) = \int \frac{p_G(Y_{-i}|Z, Y_i)}{p_G(Y_{-i}|Y_i)}p(Y_{-i}|\theta_i, Y_i)dY_{-i}$$

Locally focused approximation - with adjustment

By approximating $K(Y_i)$ by a constant, which holds e.g. when $p(Y_{-i}|\theta_i, Y_i) = p_G(Y_{-i}|Y_i)$, we once again get

$$p(\theta_i|Z) \propto p(\theta_i) \int \frac{p_G(Y_i|Z)}{p_G(Y_i)} p(Y_i|\theta_i) dY_i.$$

- Even when relaxing the assumption of $p(Y|\theta) \sim$ Gaussian this can be calculated by numerical integration since the dimension of Y_i is typically ≤ 3.

Locally focused approximation - with adjustment

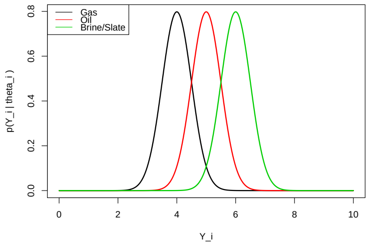
By approximating $K(Y_i)$ by a constant, which holds e.g. when $p(Y_{-i}|\theta_i, Y_i) = p_G(Y_{-i}|Y_i)$, we once again get

$$p(\theta_i|Z) \propto p(\theta_i) \int \frac{p_G(Y_i|Z)}{p_G(Y_i)} p(Y_i|\theta_i) dY_i.$$

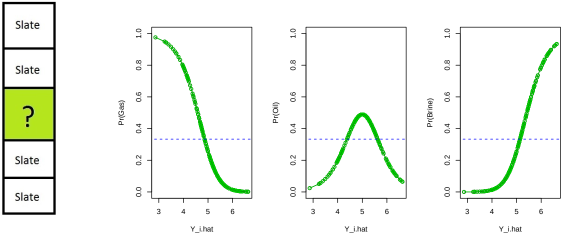
- Even when relaxing the assumption of $p(Y|\theta) \sim$ Gaussian this can be calculated by numerical integration since the dimension of Y_i is typically ≤ 3.

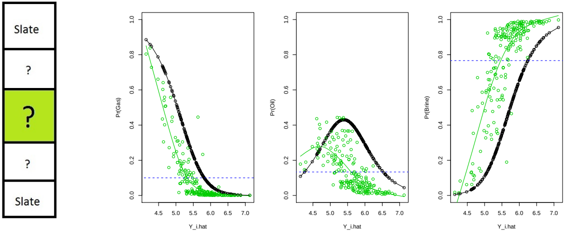
Simple example

- *One* surface location with $n = 5$ vertical positions
- $\dim(Y_i) = 1$ and Y denotes acoustic impedance
- 4 possible facies: Slate, Brine, Oil and Gas
- Ignore wavelet smoothing
- Observe $n - 1$ contrasts (Seismic data)



Two basic situations





Summary

- The approximation can be rough, but is very fast.
- Do not need to specify the whole categorical distribution $p(\theta)$, but only $p(\theta_i)$

Further work:

- Need to address in which situations $K(Y_i)$ is approximately constant
- Alternative approximations (ideas!)
  - *Nested* Laplace approximations for $p(Y_i), p(Y_i|Z)$
  - Simplified Gaussian mixture approximation for $p(Y_i), p(Y_i|Z)$
  - Wider local neighborhood approximation
  - Exact computation with simple neighborhood structure on Y.
  - Dimension reduction?
  - Other alternatives?

Summary

Klække 2013

Martin Jullum

- The approximation can be rough, but is very fast.
- Do not need to specify the whole categorical distribution $p(\theta)$, but only $p(\theta_i)$

Further work:

- Need to address in which situations $K(Y_i)$ is approximately constant
- Alternative approximations (ideas!)
  - *Nested Laplace approximations for $p(Y_i), p(Y_i|Z)$*
  - Simplified Gaussian mixture approximation for $p(Y_i), p(Y_i|Z)$
  - Wider local neighborhood approximation
  - Exact computation with simple neighborhood structure on Y
  - Dimension reduction?
  - Other alternatives?

Summary

- The approximation can be rough, but is very fast.
- Do not need to specify the whole categorical distribution $p(\theta)$, but only $p(\theta_i)$

Further work:

- Need to address in which situations $K(Y_i)$ is approximately constant
- Alternative approximations (ideas!)
  - *Nested* Laplace approximations for $p(Y_i), p(Y_i|Z)$
  - Simplified Gaussian mixture approximation for $p(Y_i), p(Y_i|Z)$
  - Wider local neighborhood approximation
  - Exact computation with simple neighborhood structure on Y
  - Dimension reduction?
  - Other alternatives?

Summary

- The approximation can be rough, but is very fast.
- Do not need to specify the whole categorical distribution $p(\theta)$, but only $p(\theta_i)$

Further work:

- Need to address in which situations $K(Y_i)$ is approximately constant
- Alternative approximations (ideas!)
  - *Nested* Laplace approximations for $p(Y_i), p(Y_i|Z)$
  - Simplified Gaussian mixture approximation for $p(Y_i), p(Y_i|Z)$
  - Wider local neighborhood approximation
  - Exact computation with simple neighborhood structure on Y.
  - Dimension reduction?
  - Other alternatives?

Summary

- The approximation can be rough, but is very fast.
- Do not need to specify the whole categorical distribution $p(\theta)$, but only $p(\theta_i)$

Further work:

- Need to address in which situations $K(Y_i)$ is approximately constant
- Alternative approximations (ideas!)
  - *Nested* Laplace approximations for $p(Y_i), p(Y_i|Z)$
  - Simplified Gaussian mixture approximation for $p(Y_i), p(Y_i|Z)$
  - Wider local neighborhood approximation
  - Exact computation with simple neighborhood structure on Y.
  - Dimension reduction?
  - Other alternatives?

Summary

- The approximation can be rough, but is very fast.
- Do not need to specify the whole categorical distribution $p(\theta)$, but only $p(\theta_i)$

Further work:

- Need to address in which situations $K(Y_i)$ is approximately constant
- Alternative approximations (ideas!)
  - *Nested* Laplace approximations for $p(Y_i), p(Y_i|Z)$
  - Simplified Gaussian mixture approximation for $p(Y_i), p(Y_i|Z)$
  - Wider local neighborhood approximation
  - Exact computation with simple neighborhood structure on Y.
  - Dimension reduction?
  - Other alternatives?

Summary

- The approximation can be rough, but is very fast.
- Do not need to specify the whole categorical distribution $p(\theta)$, but only $p(\theta_i)$

Further work:

- Need to address in which situations $K(Y_i)$ is approximately constant
- Alternative approximations (ideas!)
  - *Nested* Laplace approximations for $p(Y_i), p(Y_i|Z)$
  - Simplified Gaussian mixture approximation for $p(Y_i), p(Y_i|Z)$
  - Wider local neighborhood approximation
  - Exact computation with simple neighborhood structure on Y.
  - Dimension reduction?
  - Other alternatives?

Summary

- The approximation can be rough, but is very fast.
- Do not need to specify the whole categorical distribution $p(\theta)$, but only $p(\theta_i)$

Further work:

- Need to address in which situations $K(Y_i)$ is approximately constant
- Alternative approximations (ideas!)
  - *Nested* Laplace approximations for $p(Y_i), p(Y_i|Z)$
  - Simplified Gaussian mixture approximation for $p(Y_i), p(Y_i|Z)$
  - Wider local neighborhood approximation
  - Exact computation with simple neighborhood structure on Y.
  - Dimension reduction?
  - Other alternatives?

Summary

- The approximation can be rough, but is very fast.
- Do not need to specify the whole categorical distribution $p(\theta)$, but only $p(\theta_i)$

Further work:

- Need to address in which situations $K(Y_i)$ is approximately constant
- Alternative approximations (ideas!)
  - *Nested* Laplace approximations for $p(Y_i), p(Y_i|Z)$
  - Simplified Gaussian mixture approximation for $p(Y_i), p(Y_i|Z)$
  - Wider local neighborhood approximation
  - Exact computation with simple neighborhood structure on Y.
  - Dimension reduction?
  - Other alternatives?

Summary

- The approximation can be rough, but is very fast.
- Do not need to specify the whole categorical distribution $p(\theta)$, but only $p(\theta_i)$

Further work:

- Need to address in which situations $K(Y_i)$ is approximately constant
- Alternative approximations (ideas!)
  - *Nested* Laplace approximations for $p(Y_i), p(Y_i|Z)$
  - Simplified Gaussian mixture approximation for $p(Y_i), p(Y_i|Z)$
  - Wider local neighborhood approximation
  - Exact computation with simple neighborhood structure on Y.
  - Dimension reduction?
  - Other alternatives?

Summary

- The approximation can be rough, but is very fast.
- Do not need to specify the whole categorical distribution $p(\theta)$, but only $p(\theta_i)$

Further work:

- Need to address in which situations $K(Y_i)$ is approximately constant
- Alternative approximations (ideas!)
  - *Nested* Laplace approximations for $p(Y_i), p(Y_i|Z)$
  - Simplified Gaussian mixture approximation for $p(Y_i), p(Y_i|Z)$
  - Wider local neighborhood approximation
  - Exact computation with simple neighborhood structure on Y.
  - Dimension reduction?
  - Other alternatives?

Thank you for your time!