ECON5120 Bayesian Data Analysis

Week 7: Variational Bayes (Advanced MCMC Methods)

Santiago Montoya-Blandón

University of Glasgow
Adam Smith Business School

March 3, 2026

Lecture outline

1. Motivating Variational Bayes
2. Normal (Gaussian) Approximations
3. Variational Approximations
4. Mean-Field Approximation and Optimality
5. Variational Inference for Bayesian Linear Regression

Bayes Rule and Maximum-a-Posteriori

- Recall Bayes' rule for parameters θ with data y

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)}$$

- Marginal likelihood $p(y)$ usually intractable

Bayes Rule and Maximum-a-Posteriori

- Recall Bayes' rule for parameters θ with data y

$$p(\theta \mid y) = \frac{p(y \mid \theta)p(\theta)}{p(y)}$$

- Marginal likelihood $p(y)$ usually intractable
- What if evaluating likelihood $p(y \mid \theta)$ itself is difficult?
- Example: $\theta = (\delta, z)$ where z is a latent unobserved variable
- Joint prior for δ and z: $p(\delta, z) = p(\delta \mid z)p(z)$

$$p(y \mid \delta) = \int p(y \mid \delta, z)p(\delta \mid z)p(z)dz$$

- Likelihood needs to integrate out unobserved z and is difficult to evaluate

Bayes Rule and Maximum-a-Posteriori

- In such cases, using standard Bayesian inference is difficult
- If we cannot evaluate likelihood, using Metropolis-Hastings is not feasible
- We lose out on our most powerful algorithm so far
- How to incorporate prior information if we cannot sample from posterior?

Bayes Rule and Maximum-a-Posteriori

- In such cases, using standard Bayesian inference is difficult
- If we cannot evaluate likelihood, using Metropolis-Hastings is not feasible
- We lose out on our most powerful algorithm so far
- How to incorporate prior information if we cannot sample from posterior?
- Idea: replace *sampling step* with *optimization*
- Frequentists use MLE: $\widehat{\theta}_{\text{MLE}} = \arg\max_{\theta \in \Theta} p(y \mid \theta)$
- We could use the Maximum-a-Posteriori (MAP) estimator:

$$\widehat{\theta}_{\text{MAP}} := \arg\max_{\theta \in \Theta} p(\theta \mid y) = \arg\max_{\theta \in \Theta} \frac{p(y \mid \theta) p(\theta)}{p(y)} = \arg\max_{\theta \in \Theta} p(y \mid \theta) p(\theta)$$

Gibbs Sampling and Optimization

- As additional motivation, think of the Gibbs sampler
- Recall a 2D example: Want to sample from $p(\theta_1, \theta_2|y)$. At s-th draw:
  1. Draw $\theta_1^{(s)}$ from $p(\theta_1|\theta_2^{(s-1)}, y)$
  2. Draw $\theta_2^{(s)}$ from $p(\theta_2|\theta_1^{(s)}, y)$
- Intuitively: iterative and sequential sampling from conditional posteriors
- Recall posterior hyperparameters change with current state $\theta^{(s)}$

Gibbs Sampling and Optimization

- As additional motivation, think of the Gibbs sampler
- Recall a 2D example: Want to sample from $p(\theta_1, \theta_2|y)$. At s-th draw:
  1. Draw $\theta_1^{(s)}$ from $p(\theta_1|\theta_2^{(s-1)}, y)$
  2. Draw $\theta_2^{(s)}$ from $p(\theta_2|\theta_1^{(s)}, y)$
- Intuitively: iterative and sequential sampling from conditional posteriors
- Recall posterior hyperparameters change with current state $\theta^{(s)}$
- What if we pre-specify the posterior hyperparameters *before* sampling?
- Could we choose the pre-specified parameters optimally?
- VB idea: Choose them using *optimization* before *sampling*!

Lecture outline

1. Motivating Variational Bayes
2. Normal (Gaussian) Approximations
3. Variational Approximations
4. Mean-Field Approximation and Optimality
5. Variational Inference for Bayesian Linear Regression

Normal Approximation

- Before moving forward to VB approximations, review simplest one
- Start from a general posterior $p(\theta \mid y)$
- What if we *approximate* $p(\theta \mid y)$ as a normal?

Normal Approximation

- Before moving forward to VB approximations, review simplest one
- Start from a general posterior $p(\theta \mid y)$
- What if we *approximate* $p(\theta \mid y)$ as a normal?
- Any normal is defined by its *mean* and *variance*
- Find mean and variance using *optimization*!
- We again use Maximum-a-Posteriori (MAP), provides *posterior mode*
- Useful fact:

$$\widehat{\theta}_{\text{MAP}} = \arg\max_{\theta \in \Theta} p(\theta \mid y) \implies \widehat{\theta}_{\text{MAP}} = \arg\max_{\theta \in \Theta} [\log p(y \mid \theta) + \log p(\theta)]$$

Normal Approximation

- Choose $\widehat{\theta}_{\text{MAP}}$ as the *mean*
- Choose inverse of curvature at posterior mode as the *variance*
- Curvature defined as second derivative of a function
- Corresponds to (observed) *Fisher information* of posterior

$$I(\widehat{\theta}_{\text{MAP}}) := -\left.\frac{\partial^2 \log p(\theta \mid y)}{d\theta^2}\right|_{\theta=\widehat{\theta}_{\text{MAP}}}$$

- *Normal posterior approximation*: $\theta \mid y \underset{\text{approx}}{\sim} \mathcal{N}(\widehat{\theta}_{\text{MAP}}, [I(\widehat{\theta}_{\text{MAP}})]^{-1})$
- Intuition: choosing this approximation matches value, first and second derivative of posterior

Normal Approximation: Example

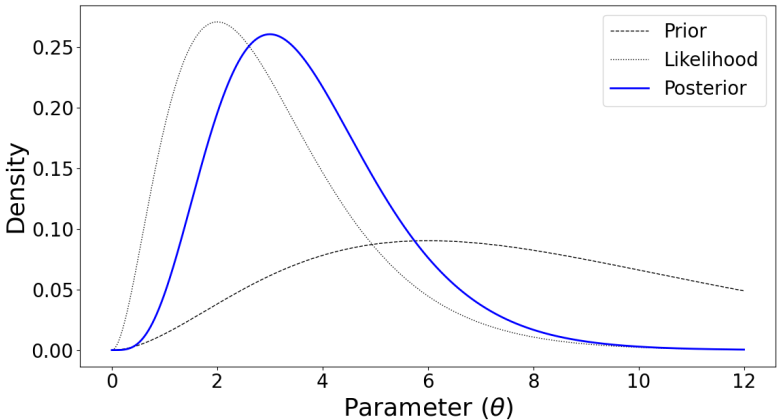

Figure: Likelihood, prior and resulting posterior

Normal Approximation: Example

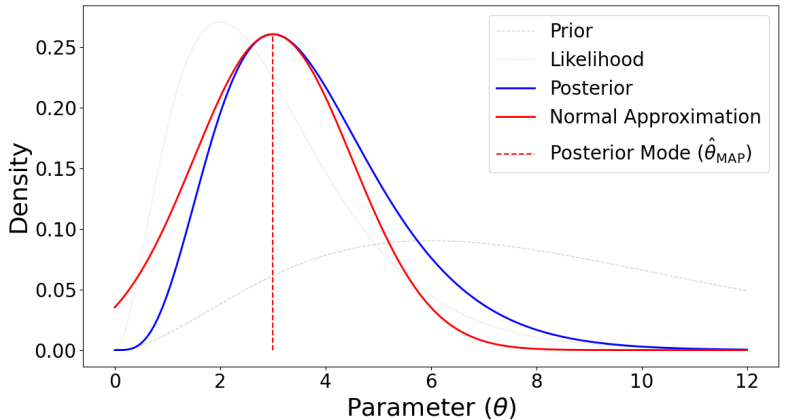

Figure: Normal approximation to posterior (posterior mode highlighted)

Normal Approximation: Details

- Denote $q(\theta) := p(y \mid \theta)p(\theta)$ (numerator of Bayes rule)
- Use a second-order Taylor expansion of $\log q(\theta)$ around $\widehat{\theta}_{\text{MAP}}$

$$\log q(\theta) \approx \underbrace{\log q(\widehat{\theta}_{\text{MAP}})}_{\text{constant w.r.t. } \theta} + (\theta - \widehat{\theta}_{\text{MAP}})^\top \underbrace{\left.\frac{\partial \log q(\theta)}{d\theta}\right|_{\theta=\widehat{\theta}_{\text{MAP}}}}_{=0 \text{ at maximum}} + \frac{1}{2}(\theta - \widehat{\theta}_{\text{MAP}})^\top \left.\frac{\partial^2 \log q(\theta)}{d\theta^2}\right|_{\theta=\widehat{\theta}_{\text{MAP}}} (\theta - \widehat{\theta}_{\text{MAP}})$$

- Transform back to $q(\theta)$ (take exp of both sides)

$$q(\theta) \approx q(\widehat{\theta}_{\text{MAP}}) \exp\left\{ -\frac{1}{2}(\theta - \widehat{\theta}_{\text{MAP}})^\top \left[-\left.\frac{\partial^2 \log q(\theta)}{d\theta^2}\right|_{\theta=\widehat{\theta}_{\text{MAP}}} \right] (\theta - \widehat{\theta}_{\text{MAP}}) \right\}$$

Normal Approximation: Details

- Recall $q(\widehat{\theta}_{\text{MAP}})$ is constant with respect to θ (of dimension d)
- To obtain full posterior, get implied normalization constant $p(y)$

$$
\begin{aligned}
p(y) &= \int p(y \mid \theta) p(\theta) d\theta \approx \int q(\theta) d\theta \\
&\approx q(\widehat{\theta}_{\text{MAP}}) \int \exp\left\{ -\frac{1}{2} (\theta - \widehat{\theta}_{\text{MAP}})^\top I(\widehat{\theta}_{\text{MAP}}) (\theta - \widehat{\theta}_{\text{MAP}}) \right\} d\theta \\
&\approx q(\widehat{\theta}_{\text{MAP}}) (2\pi)^{d/2} |I(\theta)|^{-1/2} \underbrace{\int \mathcal{N}(\theta \mid \widehat{\theta}_{\text{MAP}}, I(\theta)^{-1}) d\theta}_{=1} \\
&\approx q(\widehat{\theta}_{\text{MAP}}) (2\pi)^{d/2} |I(\theta)|^{-1/2}
\end{aligned}
$$

- Numerical approximation:

$$
p(\theta \mid y) \approx p(\widehat{\theta}_{\text{MAP}} \mid y) \exp[-0.5 (\theta - \widehat{\theta}_{\text{MAP}})^\top I(\widehat{\theta}_{\text{MAP}}) (\theta - \widehat{\theta}_{\text{MAP}})]
$$

Lecture outline

1. Motivating Variational Bayes
2. Normal (Gaussian) Approximations
3. Variational Approximations
4. Mean-Field Approximation and Optimality
5. Variational Inference for Bayesian Linear Regression

Variational Bayes

- Variational Bayes (VB) aims to *approximate* the posterior before sampling
- Approximation idea is not new, can always approximate posterior
- VB idea is to do so optimally
- By choosing our posterior approximation optimally, we get best of both worlds:
  1. Optimization: single optimization step at the start is fast
  2. Sampling: sampling is faster as posterior does not change within main loop
- Important caveat: posterior inference is now only *approximate*
- In practice: posterior locations are well estimated, posterior uncertainty not so much

Kullback-Leibler Divergence

- Question becomes: how to choose posterior approximation optimally?
- First we need a way to measure optimality
- VB focuses on the Kullback-Leibler (KL) divergence measure
- Intuition: "distance" between two distributions
- For two distributions $p(x)$ and $q(x)$, the KL divergence between them is

$$\mathrm{KL}(q||p) = \mathbb{E}_q\left[\log \frac{q(X)}{p(X)}\right] = \mathbb{E}_q[\log q(X)] - \mathbb{E}_q[\log p(X)]$$

- This measure is non-negative $KL(q||p) \geq 0$ and only 0 if $p(x) = q(x)$ are the same (*almost everywhere*; i.e., for almost all x)

Optimal Variational Approximation

- VB chooses an approximation $q(\theta)$ to $p(\theta \mid y)$ over a class of approximations $\mathcal{Q}$
- VB idea: choose q optimally to minimize KL divergence

$$q^*(\theta) := \arg \min_{q(\theta) \in \mathcal{Q}} \mathrm{KL}(q(\theta) || p(\theta \mid y))$$

Optimal Variational Approximation

- VB chooses an approximation $q(\theta)$ to $p(\theta \mid y)$ over a class of approximations $\mathcal{Q}$
- VB idea: choose q optimally to minimize KL divergence

$$q^*(\theta) := \arg \min_{q(\theta) \in \mathcal{Q}} \mathrm{KL}(q(\theta) || p(\theta \mid y))$$

- However, it is easy to see that $\mathrm{KL}(q(\theta) || p(\theta \mid y))$ depends on the intractable evidence $p(y)$

$$\begin{aligned} \mathrm{KL}(q(\theta) || p(\theta \mid y)) &= \mathbb{E}_q[\log q(\theta)] - \mathbb{E}_q[\log p(\theta \mid y)] \\ &= \mathbb{E}_q[\log q(\theta)] - \mathbb{E}_q[\log p(\theta, y)] + \log p(y) \end{aligned}$$

- How to make the optimal VB approximation $q^*(\theta)$ feasible?

KL divergence and the ELBO

- While the evidence $p(y)$ is intractable, constant with respect to θ
- Maximize the Evidence Lower Bound (ELBO) instead

$$\text{elbo}(q) = \mathbb{E}_q[\log p(\theta, y)] - \mathbb{E}_q[\log q(\theta)]$$

- ELBO defined as: $\text{elbo}(q) = -\text{KL}(q(\theta)||p(\theta \mid y)) + \log p(y)$
- The ELBO gets its name as it satisfies: $\log p(y) \geq \text{elbo}(q)$
- Re-express the ELBO: Expected likelihood minus KL divergence between prior and approximation

$$\text{elbo}(q) = \mathbb{E}_q[\log p(y \mid \theta)] - \text{KL}(q(\theta)||p(\theta))$$

- Intuition: approximation is still a compromise between likelihood and priors

General Variational Bayes

- MCMC methods draw samples $\theta^{(s)} \sim p(\theta \mid y),\ s = 1, \ldots, S$ from true posterior
- MCMC ensures draws are from posterior (after a burn-in period)
- Diagnostics help us confirm samples have converged to a stationary distribution (posterior)

General Variational Bayes

- MCMC methods draw samples $\theta^{(s)} \sim p(\theta \mid y),\ s = 1, \ldots, S$ from true posterior
- MCMC ensures draws are from posterior (after a burn-in period)
- Diagnostics help us confirm samples have converged to a stationary distribution (posterior)
- VB draws samples according to $\theta^{(s)} \sim q^*(\theta),\ s = 1, \ldots, S$
- Draws never follow true posterior, but q^* is "close" to posterior (in KL sense)
- VB methods differ according to optimal approximation $q^*(\theta)$
- We will consider a specific way to select q^*: Mean-Field Approximations

Lecture outline

1. Motivating Variational Bayes
2. Normal (Gaussian) Approximations
3. Variational Approximations
4. Mean-Field Approximation and Optimality
5. Variational Inference for Bayesian Linear Regression

Mean-Field Approximation

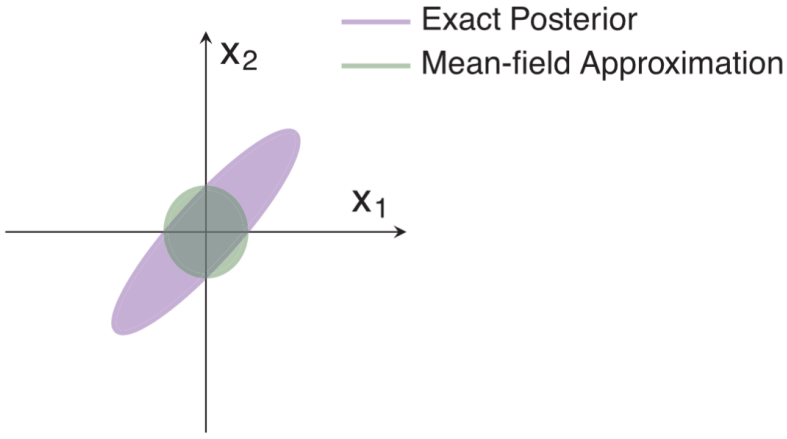

Figure: Mean-field approximation of posterior

Variational Bayes with Mean-Field Approximation

- Based on our previous discussion, VB approximation solves

$$q^*(\theta) := \arg \max_{q(\theta) \in \mathcal{Q}} \text{elbo}(q)$$

- This is still quite general, which approximation class $\mathcal{Q}$ to use?
- Standard choice: mean-field approximation. If $q \in \mathcal{Q}$, then

$$q(\theta) = \prod_{j=1}^{p} q_j(\theta_j)$$

- Intuition: approximate complicated joint posterior with product of independent distributions

Mean-Field Approximation

- The independence assumption simplifies the ELBO:

$$\begin{aligned}\text{elbo}(q) &= \mathbb{E}_q\left[\log \frac{p(\theta, y)}{q(\theta)}\right] \\ &= \mathbb{E}_q\left[\log p(y \mid \theta)\right] + \mathbb{E}_q\left[\log p(\theta)\right] - \sum_{j=1}^{p} \mathbb{E}_{q_j}\left[\log q_j(\theta_j)\right]\end{aligned}$$

- Taking expectations over an independent distribution: integrate out one parameter at a time

Optimal Mean-Field Approximation

- How to find the optimal approximation $q^*(\theta)$?
- Notation: let $\theta_{-j} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_p)$ and $E_j[f(\theta)] \coloneqq \mathbb{E}_{q_j}[f(\theta)]$
- **Coordinate Ascent**: take $q_{-j}(\theta_{-j})$ as given and optimize over $q_j(\theta_j)$

$$\text{elbo}(q_j) = \mathbb{E}_j[\mathbb{E}_{-j}[\log p(\theta \mid \theta_{-j}, y)]] - \mathbb{E}_j[\log q_j(\theta_j)] + \text{const (w.r.t. } q_j)$$

- For optimality, choose: $\log q_j^*(\theta_j) = \mathbb{E}_{-j}[\log p(\theta_j \mid \theta_{-j}, y)]$
- This results in the optimal approximations

$$q_j^*(\theta_j) = \frac{\exp\{\mathbb{E}_{-j}[\log p(\theta_j \mid \theta_{-j}, y)]\}}{\int \exp\{\mathbb{E}_{-j}[\log p(\theta_j \mid \theta_{-j}, y)]\} d\theta_j} \propto \exp\{\mathbb{E}_{-j}[\log p(\theta_j \mid \theta_{-j}, y)]\}$$

Optimal Mean-Field Approximation

- Intuition: As θ is independent under the mean-field approximation, I can integrate out $q^*_{-j}(\theta_{-j})$ for each θ_j
- Can also be used when θ is separated into blocks
- Example: in linear regression, β is p-dimensional and σ^2 is scalar
- Mean-field approximation for Bayesian linear regression:

$$q(\beta, \sigma^2) = q(\beta)q(\sigma^2)$$

- Careful: this looks like a prior, but it is an approximation to the posterior instead!
- Similar interpretation to prior though: *regularizing solutions towards* $q^*(\theta)$

Coordinate Ascent Mean-Field Variational Bayes

1. Define a model $p(y \mid \theta)$ with priors $p(\theta)$
2. Initialize the variational approximations $q_1(\theta_1), \ldots, q_p(\theta_p)$
3. For $j = 1, \ldots, p$, set $q_j(\theta_j)$ as

$$q_j(\theta_j) = \frac{\exp\{\mathbb{E}_{-j}[\log p(\theta_j \mid \theta_{-j}, y)]\}}{\int \exp\{\mathbb{E}_{-j}[\log p(\theta_j \mid \theta_{-j}, y)]\} d\theta_j}$$

4. Calculate the ELBO for updated q (either analytically or numerically)
5. Repeat steps 3 and 4 until $\text{elbo}(q)$ no longer increases
6. Sample θ from $q(\theta)$ (approximation after convergence) S times
7. Compute posterior expectations using the draws $(\theta^{(1)}, \ldots, \theta^{(S)})$

Lecture outline

1. Motivating Variational Bayes
2. Normal (Gaussian) Approximations
3. Variational Approximations
4. Mean-Field Approximation and Optimality
5. Variational Inference for Bayesian Linear Regression

Bayesian Linear Regression

- Sample of size n of outcome (y) and covariates (X)
- Focus on linear models of the form

$$y = X\beta + u, \quad u \sim \mathcal{N}(0, \sigma^2 I_n)$$

- Likelihood: $p(y \mid X; \beta, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(y - X\beta)^\top (y - X\beta)\right\}$
- Independent priors: $p(\beta, \sigma^2) = \mathcal{N}\left(\beta \mid \underline{\beta}, \underline{B}\right) \times \text{Inv-Gamma}\left(\sigma^2 \mid \underline{\alpha}/2, \underline{\delta}/2\right)$
- Gibbs sampler available, no analytical solution to posterior
- Compare with Mean-Field Variational Bayes algorithm

Mean-Field Variational Bayes Linear Regression

- Class of approximations in this case:

$$\mathcal{Q} := \{q(\beta, \sigma^2) : q(\beta, \sigma^2) = q(\beta) \cdot q(\sigma^2)\}$$

- Shapes of $q(\beta)$ and $q(\sigma^2)$ are unconstrained
- We will obtain optimal approximations given linear model
- Optimal mean-field approximation: $q^*(\theta_j) \propto \exp\{\mathbb{E}_{-j}[\log p(\theta_j \mid \theta_{-j}; y, X)]\}$
- Optimal mean-field approximations in linear case:

$$q^*(\beta) \propto \exp\{\mathbb{E}_{\sigma^2}[\log p(\beta \mid \sigma^2; y, X)]\}$$
$$q^*(\sigma^2) \propto \exp\{\mathbb{E}_{\beta}[\log p(\sigma^2 \mid \beta; y, X)]\}$$

- Only need expectation over conditional posteriors (already derived!)

Optimal Approximations in Linear Regression

Optimal approximation for coefficients:

$$\log q^*(\beta) = \mathbb{E}_{q^*(\sigma^2)}[\log p(\beta \mid \sigma^2; y, X)] = -\frac{1}{2}(\beta - \bar{\beta})^\top \bar{B}^{-1}(\beta - \bar{\beta}) + \text{const.}$$

$$\bar{B} := \left(\mathbb{E}_{q^*(\sigma^2)} \left[\frac{1}{\sigma^2} \right] X^\top X + \underline{B}^{-1} \right)^{-1}$$

$$\bar{\beta} := \bar{B} \left(\mathbb{E}_{q^*(\sigma^2)} \left[\frac{1}{\sigma^2} \right] X^\top y + \underline{B}^{-1} \underline{\beta} \right)$$

Optimal approximation for variance:

$$\log q^*(\sigma^2) = \mathbb{E}_{q^*(\beta)}[\log p(\sigma^2 \mid \beta; y, X)] = -\frac{\bar{\delta}}{2\sigma^2} - \left(\frac{\bar{\alpha}}{2} + 1 \right) \log(\sigma^2) + \text{const.}$$

$$\bar{\alpha} := \underline{\alpha} + n$$

$$\bar{\delta} := \underline{\delta} + \mathbb{E}_{q^*(\beta)} \left[(y - X\beta)^\top (y - X\beta) \right]$$

Optimal Approximations in Linear Regression

- Optimal approximation for coefficients:

$$q^*(\beta) \propto \exp\left\{-\frac{1}{2}(\beta-\bar{\beta})^\top \bar{B}^{-1}(\beta-\bar{\beta})\right\} \implies q^*(\beta) = \mathcal{N}(\beta \mid \bar{\beta}, \bar{B})$$

- Optimal approximation for variance:

$$q^*(\sigma^2) \propto (\sigma^2)^{-\bar{\alpha}/2-1} \exp\left\{-\frac{\bar{\delta}}{2\sigma^2}\right\} \implies q^*(\sigma^2) = \text{Inv-Gamma}\left(\sigma^2 \middle| \frac{\bar{\alpha}}{2}, \frac{\bar{\delta}}{2}\right)$$

- To make operational, use following known results:

$$\mathbb{E}_{q^*(\sigma^2)}\left[\frac{1}{\sigma^2}\right] = \frac{\bar{\alpha}}{\bar{\delta}}$$

$$\mathbb{E}_{q^*(\beta)}\left[(y-X\beta)^\top(y-X\beta)\right] = \text{tr}(\bar{B}X^\top X) + (y-X\bar{\beta})^\top(y-X\bar{\beta})$$

Mean-Field Variational Bayes Linear Regression Algorithm

1. Start from an initial value for $(\bar{\beta}, \bar{B}, \bar{\alpha}, \bar{\delta})$
2. Update $(\bar{\beta}, \bar{B}, \bar{\alpha}, \bar{\delta})$ to a new value $(\tilde{\beta}, \tilde{B}, \tilde{\alpha}, \tilde{\delta})$ using optimal approximations
3. Calculate $elbo(q^*)$ of the optimal approximation
4. Repeat step 2 until $elbo(q^*)$ values no longer change
5. Collect final values as $(\beta^*, B^*, \alpha^*, \delta^*)$
6. Sample S draws from $\beta^{(s)} \sim \mathcal{N}(\beta^*, B^*)$
7. Sample S draws from $\sigma^{2(s)} \sim \text{Inv-Gamma}(\alpha^*/2, \delta^*/2)$
8. Compute posterior expectations using draws $\{\beta^{(s)}, \sigma^{2(s)}\}_{s=1}^{S}$

Variational Bayes for Linear Regression

- Computation of $elbo(q^*)$ can be done in closed-form for linear model
- Another alternative: update hyperparameters until they converge
- In practice, VB provides a *fast* and *accurate* approximation to posterior expectation estimates
- Caveat: uncertainty captured by VB usually less than true posterior
- Point estimates are generally correct but credibility intervals are too narrow

Variational Bayes and Gibbs Sampling

Mean-field optimal approximations

$$q^*(\beta) = \mathcal{N}(\beta \mid \bar{\beta}, \bar{B})$$

$$q^*(\sigma^2) = \text{Inv-Gamma}\left(\sigma^2 \middle| \frac{\bar{\alpha}}{2}, \frac{\bar{\delta}}{2}\right)$$

with hyperparameters

$$\bar{B} := \left(\mathbb{E}_{q^*(\sigma^2)}\left[\frac{1}{\sigma^2}\right] X^\top X + \underline{B}^{-1}\right)^{-1}$$

$$\bar{\beta} := \bar{B}\left(\mathbb{E}_{q^*(\sigma^2)}\left[\frac{1}{\sigma^2}\right] X^\top y + \underline{B}^{-1}\underline{\beta}\right)$$

$$\bar{\alpha} := \underline{\alpha} + n$$

$$\bar{\delta} := \underline{\delta} + \mathbb{E}_{q^*(\beta)}\left[(y - X\beta)^\top (y - X\beta)\right]$$

Gibbs sampling conditional posteriors

$$p(\beta \mid \sigma^2; y, X) = \mathcal{N}(\beta \mid \tilde{\beta}, \tilde{B})$$

$$p(\sigma^2 \mid \beta; y, X) = \text{Inv-Gamma}\left(\sigma^2 \middle| \frac{\tilde{\alpha}}{2}, \frac{\tilde{\delta}}{2}\right)$$

with hyperparameters

$$\tilde{B} := \left(\frac{1}{\sigma^2} X^\top X + \underline{B}^{-1}\right)^{-1}$$

$$\tilde{\beta} := \tilde{B}\left(\frac{1}{\sigma^2} X^\top y + \underline{B}^{-1}\underline{\beta}\right)$$

$$\tilde{\alpha} := \underline{\alpha} + n$$

$$\tilde{\delta} := \underline{\delta} + (y - X\beta)^\top (y - X\beta)$$