Mean field variational inference

Mean field variational inference is a variational inference algorithm built on the assertion that the variational family is fully factorized. That is, given latent random variables $Z := \{Z_1, \ldots, Z_n\}$ and observed random variables $X := \{X_1, \ldots, X_n\}$, we seek an approximation to the posterior of the form

$$p(z \mid x) \approx \prod_{i=1}^{n} q_i(z_i)$$

where $z := \{z_1, \ldots, z_n\}$ represents an assignment to the set of latent random variables, $x := \{x_1, \ldots, x_n\}$ represents the value of the observed random variables, and $q_i(z_i)$ is the density/mass function for the variational distribution of $Z_i$. We note that this assumption does not require any assumed specific form for each $q_i(z_i)$.

Description

Recall, variational inference attempts to find $q$ that maximizes the evidence lower bound (ELBO), given by

$$\text{ELBO}(q) = \int q(z) \log p(x, z) - q(z) \log q(z) \, dz$$

When we assert that the variational family has the aforementioned factorized form, we can derive a coordinate ascent algorithm. If, for a given $j$, we fix the distributions for $i \neq j$ to $q_i \neq j$, we can show that the $q_j$ that maximizes the ELBO is given by:

$$q_j(z_j) = \frac{\exp(E_{z_i \neq j \sim q_{\neq j}} [\log p(x, z_{\neq j}, z_j)])}{\int \exp(E_{z_i \neq j \sim q_{\neq j}} [\log p(x, z_{\neq j}, z_j)]) \, dz_j}$$ (1)

where $p(x, z_{\neq j}, z_j)$ is the probability density/mass function $p(x, z)$ rewritten so that the random variables $z_i$ for $i \neq j$ are grouped together separately from $z_j$ (Theorem 1). Thus, the coordinate ascent algorithm involves cycling through each $q_j$, and updating $q_j$ according to right-hand side of Equation 1.

However, computing the denominator of the right-hand side of Equation 1 may not be tractable. We can therefore modify the coordinate ascent algorithm to update the unnormalized $q_j$ for each $j \in \{1, \ldots, n\}$. That is, we realize that

$$q_j \propto \exp(E_{z_{\neq j} \sim q_{\neq j}} [\log p(z_j \mid x, z_{\neq j}, z_j)])$$ (2)

(Theorem 2) and thus, we can simply compute this term on each iteration. When devising a mean field variational inference algorithm, the challenge amounts to computing the
Algorithm 1 Mean field variational inference

Precondition:
1. A joint probabilistic model $p(x, z)$ over random variables $Z := \{Z_1, \ldots, Z_n\}$ and $X := \{X_1, \ldots, X_n\}$ where $z := \{z_1, \ldots, z_n\}$ represents an assignment to the set of latent random variables, $x := \{x_1, \ldots, x_n\}$ represents the value of the observed random variables

2. $\tilde{q}_1, \ldots, \tilde{q}_n \leftarrow$ Unnormalized density/mass functions for any set of distributions over $Z_1, \ldots, Z_n$
3. while $\tilde{q}_1, \ldots, \tilde{q}_n$ have not converged do
   4. for $j = 1, \ldots, n$ do
      5. $\tilde{q}_j \leftarrow \exp \left( E_{\tilde{q}_{\neq j} \rightarrow \tilde{q}_j} \left[ \log p(z_j \mid x, z_{\neq j}, z_j) \right] \right)$ \hspace{1cm} \text{Compute the unnormalized density/mass function for} \hspace{1cm} q_j$
   6. end for
4. end while
5. return $\tilde{q}_1, \ldots, \tilde{q}_n$

expectation on the right-hand side of Equation 2. Mean field variational inference is therefore usually only tractable when the full joint model $p(x, z)$ yields a nice closed form solution to Equation 2, which is not guaranteed for every model.

Theorem 1 Given a set of latent random variables $Z := \{Z_1, \ldots, Z_n\}$, a set of observed random variables $X := \{X_1, \ldots, X_n\}$, and a factorized distribution of the form

$$q(z) := \prod_{i=1}^{n} q_i$$

where $q_i := q_i(z_i)$, for some $q_j$, if we fix the remaining distributions $q_{\neq j}$,

$$\hat{q}_j = \underset{q_j}{\text{argmax}} \text{ELBO}(q_j) \implies \hat{q}_j(z_j) = \frac{\exp \left( E_{\tilde{q}_{\neq j} \rightarrow \tilde{q}_j} \left[ \log p(x, z_{\neq j}, z_j) \right] \right)}{\int \exp \left( E_{\tilde{q}_{\neq j} \rightarrow \tilde{q}_j} \left[ \log p(x, z_{\neq j}, z_j) \right] \right) \, dz_j}$$

where

$$\text{ELBO}(q_j) = \int \left[ q_j \prod_{i \neq j} q_i \right] \left[ \log p(x, z) - \left( \log q_j + \sum_{i \neq j} \log q_i \right) \right] \, dz \quad (3)$$
Proof:

\[
\text{ELBO}(q_j) = \int \left[ q_j \prod_{i \neq j} q_i \right] \left[ \log p(x, z) - (\log q_j + \sum_{i \neq j} \log q_i) \right] \, dz
\]

\[
= \int q_j \int_{z_{i \neq j}} \left[ \prod_{i \neq j} q_i \right] \left[ \log p(x, z) - (\log q_j + \sum_{i \neq j} \log q_i) \right] \, dz_{i \neq j} \, dz_j
\]

\[
= \int_{z_j} q_j \int_{z_{i \neq j}} \left[ \prod_{i \neq j} q_i \right] \left[ \log p(x, z) \right] \, dz_{i \neq j} \, dz_j
\]

\[
- \int_{z_j} q_j \int_{z_{i \neq j}} \left[ \prod_{i \neq j} q_i \right] \left[ \log q_j + \sum_{i \neq j} \log q_i \right] \, dz_{i \neq j} \, dz_j
\]

\[
= \int_{z_j} q_j \log \tilde{p}(x, z_j) \, dz_j + \int_{z_j} q_j \log q_j \, dz_j + C
\]

\[
\text{See Note 1} + \text{See note 2}
\]

\[
= -KL(\tilde{p}(x, z_j) || q_j(z_j)
\]

where C is a constant and

\[
\log \tilde{p}(x, z_j) := E_{z_{i \neq j}} q_{i \neq j} \left[ \log p(x, z_{i \neq j}, z_j) \right] + K
\]

where K is a constant. Then, it is clear that

\[
\hat{q}_j := \tilde{p}(x, z_j)
\]

maximizes ELBO(q_j) because it causes \(KL(\tilde{p}(x, z_j) || q_j(z_j))\) to be zero. Rewriting \(\hat{q}_j\), we have

\[
\log \hat{q}_j = \log \tilde{p}(x, z_j)
\]

\[
= E_{z_{i \neq j}} q_{i \neq j} \left[ \log p(x, z_{i \neq j}, z_j) \right] + K
\]

\[
\Rightarrow \hat{q}_j = \exp \left( E_{z_{i \neq j}} q_{i \neq j} \left[ \log p(x, z_{i \neq j}, z_j) \right] + K \right)
\]

\[
= \exp \left( E_{z_{i \neq j}} q_{i \neq j} \left[ \log p(x, z_{i \neq j}, z_j) \right] \right) \exp(K)
\]

\[
= \frac{\exp \left( E_{z_{i \neq j}} q_{i \neq j} \left[ \log p(x, z_{i \neq j}, z_j) \right] \right)}{\int \exp \left( E_{z_{i \neq j}} q_{i \neq j} \left[ \log p(x, z_{i \neq j}, z_j) \right] \right) \, dz_j}
\]

The constant \(\exp K\) is the normalization constant.

Notes
1. 

\[ \int q_j \int_{z_{i,j}} \left[ \prod_{i \neq j} q_i \right] \left[ \log p(x, z) \right] dz_{i\neq j} dz_j \]
\[ = \int q_j E_{z_{i,j} \sim q_{i,j}} \left[ \log p(x, z_{i\neq j}, z_j) \right] dz_j \]
\[ = \int z_j \log \tilde{p}(x, z_j) - K dz_j \]
\[ = \int z_j \log \tilde{p}(x, z_j) dz_j - \int K dz_j \]
\[ = \int z_j \log \tilde{p}(x, z_j) dz_j + K' \]

The \( K' \) constant is absorbed into the constant \( C \).

2. 

\[ \int q_j \int_{z_{i,j}} \left[ \prod_{i \neq j} q_i \right] \log q_j + \sum_{i \neq j} \log q_i dz_{i\neq j} dz_j \]
\[ = \int q_j \int_{z_{i,j}} \left[ \prod_{i \neq j} q_i \right] \log q_j dz_{i\neq j} dz_j + \int q_j \int_{z_{i,j}} \left[ \prod_{i \neq j} q_i \right] \sum_{i \neq j} \log q_i dz_{i\neq j} dz_j \]
\[ = \int q_j \int_{z_{i,j}} \left[ \prod_{i \neq j} q_i \right] \log q_j dz_{i\neq j} dz_j + \hat{K} \int_{z_j} q_j dz_j \]

integrates to 1

\[ = \int q_j \int_{z_{i,j}} \left[ \prod_{i \neq j} q_i \right] \log q_j dz_{i\neq j} dz_j + \hat{K} \]
\[ = \int q_j \log q_j \int_{z_{i,j}} \left[ \prod_{i \neq j} q_i \right] dz_{i\neq j} dz_j + \hat{K} \]

integrates to 1

\[ = \int q_j \log q_j dz_j + \hat{K} \]

where

\[ \hat{K} := \int_{z_{i,j}} \left[ \prod_{i \neq j} q_i \right] \sum_{i \neq j} \log q_i dz_{i\neq j} \]
is a constant with respect to $q_j$. This constant is absorbed into $C$.

**Theorem 2**

$$q_j \propto \exp\left( E_{z_i \neq j} \log p(x, z_{i\neq j}, z_j) \right)$$

**Proof:**

$$q_j \propto \exp\left( E_{z_i \neq j} \log p(x, z_{i\neq j}, z_j) \right)$$

$$= \exp\left( E_{z_i \neq j} \log\left[p(z_j | x, z_{i\neq j}, z_j) p(x, z_{i\neq j})\right]\right)$$

$$= \exp\left( E_{z_i \neq j} \log p(z_j | x, z_{i\neq j}, z_j) + \log p(x, z_{i\neq j})\right)$$

$$= \exp\left( E_{z_i \neq j} \log p(z_j | x, z_{i\neq j}, z_j) \right) + E_{z_i \neq j} \log p(x, z_{i\neq j})$$

$$= \exp\left( E_{z_i \neq j} \log p(z_j | x, z_{i\neq j}, z_j) \right) \exp\left( E_{z_i \neq j} \log p(x, z_{i\neq j})\right)$$

$$\propto \exp\left( E_{z_i \neq j} \log p(z_j | x, z_{i\neq j}, z_j) \right)$$