# 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 $Z := \{Z_1, \dots, Z_n\}$ and observed random variables $X := \{X_1, \dots, X_n\}$, we seek an approximation to the posterior of the form

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

where $z := \{z_1, \dots, z_n\}$ represents an assignment to the set of latent random variables, $x := \{x_1, \dots, x_n\}$ 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:

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

where $p(x, z_{i \neq j}, z_j)$ 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 $j \in \{1, \dots, n\}$. That is, we realize that

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

(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:**

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

$\tilde{q}_1, \dots, \tilde{q}_n \leftarrow$ Unnormalized density/mass functions for any set of distributions over $Z_1, \dots, Z_n$
**while** $\tilde{q}_1, \dots, \tilde{q}_n$ have not converged **do**
    **for** j = 1, …, n **do**
        $\tilde{q}_j \leftarrow \exp\left(E_{z_{i \neq j} \sim q_{i \neq j}}\left[\log p(z_j \mid x, z_{i \neq j}, z_j)\right]\right)$ ▷ Compute the unnormalized density/mass function for $q_j$
    **end for**
**end while**
**return** $\tilde{q}_1, \dots, \tilde{q}_n$

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 $Z := \{Z_1, \dots, Z_n\}$, a set of observed random variables $X := \{X_1, \dots, X_n\}$, 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 $q_{i \neq j}$,*

$$\hat{q}_j = argmax_{q_j} ELBO(q_j) \implies \hat{q}_j(z_j) = \frac{\exp\left(E_{z_{i \neq j} \sim q_{i \neq j}}\left[\log p(x, z_{i \neq j}, z_j)\right]\right)}{\int \exp\left(E_{z_{i \neq j} \sim q_{i \neq j}}\left[\log p(x, z_{i \neq j}, z_j)\right]\right) dz_j}$$

*where*

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

**Proof:**

$$\begin{aligned}
\text{ELBO}(q_j) &= \int \left[q_j \prod_{i \neq j} q_i\right]\left[\log p(x,z) - (\log q_j + \sum_{i \neq j} \log q_i)\right] dz \\
&= \int_{z_j} q_j \int_{z_{i\neq j}} \left[\prod_{i \neq j} q_i\right]\left[\log p(x,z) - (\log q_j + \sum_{i \neq j} \log q_i)\right] dz_{i\neq j} dz_j \\
&= \int_{z_j} q_j \int_{z_{i\neq j}} \left[\prod_{i \neq j} q_i\right][\log p(x,z)] \; dz_{i\neq j} dz_j \\
&\qquad - \int_{z_j} q_j \int_{z_{i\neq j}} \left[\prod_{i \neq j} q_i\right]\left[\log q_j + \sum_{i \neq j} \log q_i\right] dz_{i\neq j} dz_j \\
&= \underbrace{\int_{z_j} q_j \log \tilde{p}(x, z_j) dz_j}_{\text{See Note 1}} + \underbrace{\int_{z_j} q_j \log q_j dz_j}_{\text{See note 2}} + C \\
&= -KL(\tilde{p}(x, z_j) \parallel q_j(z_j
\end{aligned}$$

where $C$ is a constant and

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

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 $KL(\tilde{p}(x, z_j) \parallel q_j(z_j))$ to be zero. Rewriting $\hat{q}_j$, we have

$$\begin{aligned}
\log \hat{q}_j &= \log \tilde{p}(x, z_j) \\
&= E_{z_{i\neq j} \sim q_i} [\log p(x, z_{i\neq j}, z_j)] + K \\
\implies \hat{q}_j &= \exp\left(E_{z_{i\neq j} \sim q_{i\neq j}} [\log p(x, z_{i\neq j}, z_j)] + K\right) \\
&= \exp\left(E_{z_{i\neq j} \sim q_{i\neq j}} [\log p(x, z_{i\neq j}, z_j)]\right) \exp(K) \\
&= \frac{\exp\left(E_{z_{i\neq j} \sim q_{i\neq j}} [\log p(x, z_{i\neq j}, z_j)]\right)}{\int \exp\left(E_{z_{i\neq j} \sim q_{i\neq j}} [\log p(x, z_{i\neq j}, z_j)]\right) \, dz_j}
\end{aligned}$$

The constant $\exp K$ is the normalization constant.

**Notes**

1.

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

The $K'$ constant is absorbed into the constant $C$.

2.

$$\begin{aligned}
&\int_{z_j} q_j \int_{z_{i\neq j}} \left[\prod_{i\neq j} q_i\right] \left[\log q_j + \sum_{i\neq j} \log q_i\right] \; dz_{i\neq j} dz_j \\
&= \int_{z_j} q_j \int_{z_{i\neq j}} \left[\prod_{i\neq j} q_i\right] \log q_j \; dz_{i\neq j} dz_j + \int_{z_j} q_j \int_{z_{i\neq j}} \left[\prod_{i\neq j} q_i\right] \sum_{i\neq j} \log q_i \; dz_{i\neq j} dz_j \\
&= \int_{z_j} q_j \int_{z_{i\neq j}} \left[\prod_{i\neq j} q_i\right] \log q_j \; dz_{i\neq j} dz_j + \hat{K} \underbrace{\int_{z_j} q_j \; dz_j}_{\text{integrates to 1}} \\
&= \int_{z_j} q_j \int_{z_{i\neq j}} \left[\prod_{i\neq j} q_i\right] \log q_j \; dz_{i\neq j} dz_j + \hat{K} \\
&= \int_{z_j} q_j \log q_j \underbrace{\int_{z_{i\neq j}} \prod_{i\neq j} q_i \; dz_{i\neq j}}_{\text{integrates to 1}} dz_j + \hat{K} \\
&= \int_{z_j} q_j \log q_j dz_j + \hat{K}
\end{aligned}$$

where

$$\hat{K} := \int_{z_{i\neq j}} \left[\prod_{i\neq j} q_i\right] \sum_{i\neq j} \log q_i \; dz_{i\neq j}$$

is a constant with respect to $q_j$. This constant is absorbed into $C$.

**Theorem 2**

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

**Proof:**

$$\begin{aligned}
q_j &\propto \exp\left(E_{z_{i \neq j} \sim q_{i \neq j}}\left[\log p(x, z_{i \neq j}, z_j)\right]\right) \\
&= \exp\left(E_{z_{i \neq j} \sim q_{i \neq j}}\left[\log[p(z_j \mid x, z_{i \neq j}, z_j) p(x, z_{i \neq j})]\right]\right) \\
&= \exp\left(E_{z_{i \neq j} \sim q_{i \neq j}}\left[\log p(z_j \mid x, z_{i \neq j}, z_j) + \log p(x, z_{i \neq j})\right]\right) \\
&= \exp\left(E_{z_{i \neq j} \sim q_{i \neq j}}\left[\log p(z_j \mid x, z_{i \neq j}, z_j)\right] + E_{z_{i \neq j} \sim q_{i \neq j}}\left[\log p(x, z_{i \neq j})\right]\right) \\
&= \exp\left(E_{z_{i \neq j} \sim q_{i \neq j}}\left[\log p(z_j \mid x, z_{i \neq j}, z_j)\right]\right) \exp\left(E_{z_{i \neq j} \sim q_{i \neq j}}\left[\log p(x, z_{i \neq j})\right]\right) \\
&\propto \exp\left(E_{z_{i \neq j} \sim q_{i \neq j}}\left[\log p(z_j \mid x, z_{i \neq j}, z_j)\right]\right)
\end{aligned}$$