# Variational Inference

Blei et al. 2017

May 12, 2021

**Summary**

The authors discuss variational inference (VI), a method that approximates an intractable function by trying to minimize the Kullback-Leibler divergence (KL) between the function itself and a simpler family of functions.

**Keywords:** algorithms; statistical computing; computationally intensive methods; variational inference; Kullback-Leibler divergence.

## 1 Introduction

Instead of using MCMC algorithms to compute the posterior $p(\boldsymbol{z}|\boldsymbol{x})$, we can set a family of approximate densities $\mathscr{D}$ over the latent variables $\boldsymbol{z}$ and find the member such that

$$q^*(\boldsymbol{z}) = \operatorname*{argmin}_{q \in \mathscr{Q}} \mathrm{KL}(q(\boldsymbol{z})||p(\boldsymbol{z}|\boldsymbol{x})) \tag{1}$$

### Key ideas

1. VI turns *inference* into *optimization*.
2. $\mathscr{D}$ has to be flexible yet efficiently optimizable.

### Considerations

› VI should be used when inference can be imprecise and we have to deal with large scale models.

› It can be used also when MCMC is not an option even for small datasets.

## 2 Variational Inference

### 2.1 Evidence Lower BOund (ELBO)

The KL divergence in (1) can be written as

$$\begin{aligned}\mathrm{KL}\big(q(\boldsymbol{z})||p(\boldsymbol{z}|\boldsymbol{x})\big) &= \mathbb{E}_q\left[\log q(\boldsymbol{z})\right] - \mathbb{E}_q\left[\log p(\boldsymbol{z}|\boldsymbol{x})\right]\\ &= \mathbb{E}_q\left[\log(q(\boldsymbol{z}))\right] - \mathbb{E}_q\left[\log p(\boldsymbol{z},\boldsymbol{x})\right] + \log p(\boldsymbol{x})\end{aligned}$$

which is still problematic since we need to compute the normalizing constant $p(\boldsymbol{x})$. Therefore, we maximize instead the *Evidence Lower BOund* (ELBO) by discarding this term

$$\mathrm{ELBO}(q) = \mathbb{E}_q\left[\log p(\boldsymbol{z},\boldsymbol{x})\right] - \mathbb{E}_q\left[\log q(\boldsymbol{z})\right], \tag{2}$$

which has the property of $\mathrm{ELBO}(q) \leq \log p(\boldsymbol{x})$ and is used as a proxy on the premise that it is a good approximation to the marginal likelihood of $\boldsymbol{x}$. However, even though this sometimes work, it's not theoretically justified.

### 2.2 Mean-field variational family

The *mean-field variational family* is a family $\mathscr{Q}$ such that

$$q(\boldsymbol{z}) = \prod_{j=1}^{m} q_j(z_j), \tag{3}$$

which can be a product of different densities (all Gaussian, some Gaussian and some categorical, ...).

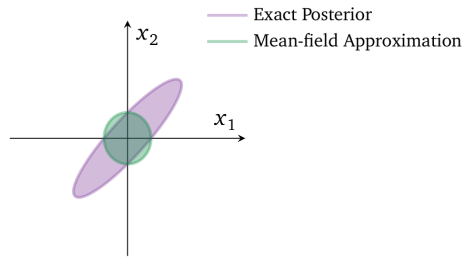


Figure 1: Mean-field approximation of a two-dimensional Gaussian posterior.

### 2.3 Coordinate ascent optimization

Coordinate ascent can be used to optimize the ELBO componentwise to a local optimum.

1. The optimal conditional distribution for $z_j$, $p(z_j|\boldsymbol{z}_{-j}, \boldsymbol{x})$, is proportional to

$$\begin{aligned} q_j^*(z_j) &\propto \exp\left\{\mathbb{E}_{-j}\left[\log p(z_j|\boldsymbol{z}_{-j}, \boldsymbol{x})\right]\right\} \\ &\propto \exp\left\{\mathbb{E}_{-j}\left[\log p(z_j, \boldsymbol{z}_{-j}, \boldsymbol{x})\right]\right\}, \end{aligned}$$

   and since the $z_j$'s are independent, the expectation does not involve $z_j$.

2. 1) is applied iteratively until convergence of the ELBO.

**Log-sum-exp trick** When dealing with logarithms of probabilities, the "log-sum-exp" trick is useful for numerical stability,

$$\log\left\{\sum_j \exp x_j\right\} = a + \log\left\{\sum_j \exp(x_j - a)\right\},$$

with $a$ usually set to $\max_i x_i$.

**Example (Various examples of variational inference)**
Example in §3 of the paper.

## 3 Variational inference for exponential families

**TODO** later