Variational inference is widely used to approximate posterior densities for Bayesian models, an alternative strategy to Markov chain Monte Carlo (MCMC) sampling. Compared to MCMC, variational inference tends to be faster and easier to scale to large data. Rather than use sampling, the main idea behind variational inference is to use optimization. Consider a joint density of latent variable $\mathbf{z} = z_{1:m}$ and observations $\mathbf{x} = x_{1:n}$,

$$p(\mathbf{z}, \mathbf{x}) = p(\mathbf{z})p(\mathbf{x}|\mathbf{z}).$$

A Bayesian model draws the latent variables from a prior density $p(\mathbf{z})$ and then relates them to the observations through the likelihood $p(\mathbf{x} \mid \mathbf{z})$. Inference in Bayesian model amounts to conditioning on data and computing the posterior $p(\mathbf{z}|\mathbf{x})$. In variational inference, we posit a family of approximate densities $\mathcal{D}$, which is a set of densities over the latent variables. Then, we try to find the member of that family that minimizes the Kullback-Leibler (KL) divergence to the exact posterior,

$$q^*(\mathbf{z}) = \underset{q(\mathbf{z}\in\mathcal{D})}{\arg\min}\ \mathrm{KL}(q(\mathbf{z})||p(\mathbf{z}|\mathbf{x})),$$

which is equivalent to maximizing the evidence lower bound (ELBO),

$$\mathrm{ELBO}(q) = \mathrm{E}[\log\ p(\mathbf{z}, \mathbf{x})] - \mathrm{E}[\log\ q(\mathbf{z})]. \tag{1}$$

Finally, we approximate the posterior with the optimized member of the family $q^*(\cdot)$.
The complexity of the family determines the complexity of the optimization; it is more difficult to optimize over a complex family than a simple family. In mean-field variational inference, we assume the latent variables are mutually independent and each governed by a distinct factor in the variational density. A generic member of the mean-field variational family is

$$q(\mathbf{z}) = \prod_{j=1}^{m} q_j(z_j).$$

Each latent variable $z_j$ is governed by its own variational factor, the density $q_j(z_j)$. In optimization, these variational factors are chosen to maximize the ELBO of (1). To solve this optimization problem, one of the most commonly used algorithms is coordinate ascent variational inference (CAVI). Consider the $j$th latent variable $z_j$. The complete conditional of $z_j$ is its conditional density given all of the other latent variables in the model and the observations, $p(z_j|\mathbf{z}_{-j}, \mathbf{x})$. Fix the other variational factor $q_l(z_l)$, $l \neq j$. The optimal $q_j(z_j)$ is then proportional to the exponentiated expected log of the complete conditional,

$$q_j^*(z_j) \propto \exp\{\mathrm{E}_{-j}[\log\ p(z_j|\mathbf{z}_{-j}, \mathbf{x})].\} \tag{2}$$

The expectation in (2) is with respect to the (currently fixed) variational density over $\mathbf{z}_{-j}$, that is $\prod_{l\neq j} q_l(z_l)$. Equivalently, (2) is proportional to the exponentiated log of the joint,

$$q_j^*(z_j) \propto \exp\{\mathrm{E}_{-j}[\log\ p(z_j, \mathbf{z}_{-j}, \mathbf{x})].\} \tag{3}$$

Because of the mean-field family assumption–that all the latent variables are independent.
Though it is flexible, the mean-field family makes strong independence assumptions. These assumptions help with scalable optimization, but they limit the expressibility of the variational family. Further, they can exacerbate issues with local optima of the objective and underestimating posterior variance. Thus, many researchers are trying to develop better approximation while maintaining efficient optimization, such as semi-implicit variational inference (SIVI) and hierarchical implicit models (HIMs).