

Black Box Variational Inference (Ranganath et al., 2014)

Satvik Saha (ss7228)

September 16, 2025

“Black box”¹ variational inference solves the problem of optimizing the Evidence Lower Bound (ELBO) $\mathcal{L}$ (over a variational family $\{q_\lambda\}$) by ‘simply’ performing (stochastic) gradient descent on the parameter λ . This paper shows that VI is possible with access only to the model joint density $p(x, z)$, variational densities q_λ (with scores $\nabla_\lambda \log q_\lambda$), and samples from q_λ : these combine to give a Monte-Carlo estimate of $\nabla_\lambda \mathcal{L}$, which informs the direction in which λ moves. A couple of variance reduction techniques (Rao-Blackwellization and control variates) are also suggested to counter the noisiness of the estimate of $\nabla_\lambda \mathcal{L}$.

In hindsight, it may seem surprising that such a direct approach (optimize $\mathcal{L}(\lambda)$ directly via gradient descent) took time to gain traction. I would guess that Monte-Carlo estimation of $\nabla_\lambda \mathcal{L}$ was seen as too noisy to be practical at the time. The variance reduction ideas in this paper are therefore crucial to its viability; the authors point out that earlier related work had struggled with this issue. Another factor may have been the field’s emphasis on analytical, model-specific updates (as in CAVI), in contrast with BBVI’s generic stochastic optimization approach. I would be curious to learn more about how this paper was initially received, given that today such methods feel almost indispensable.

¹Prof. Blei has suggested in his book that “turnkey” better captures the intent: not mysterious, just mechanical. The rise of tooling that hides the details from practitioners may explain why “black box” stuck.