Semi Implicit Variational Inference for Bayesian Spatial Models

Patrick Ding

5/9/2019

Bayesian inference

Given a (spatial) model $p(x|z)$ and prior $p(z)$, want to find posterior

$$p(z|x) = \frac{p(x|z)p(z)}{\int p(x|z)p(z)dz}$$

Often denominator intractable.

MCMC

Popular solution since late 80's: Markov chain Monte Carlo (MCMC) methods.

Create Markov chain with stationary distribution the same as the desired posterior distribution. Step along the chain to draw samples.

Pros and cons

Pros:

- Markov chain reaches the stationary distribution (posterior) eventually.

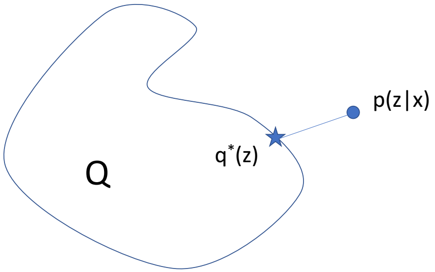
Cons:

- Can be slow (not as readily parallelizable).

Variational inference

Approximate the intractable posterior with a distribution $q(z) \in \mathcal{Q}$, where $\mathcal{Q}$ is a family of tractable distributions.

$$\min_{q(z) \in \mathcal{Q}} KL(q(z)||p(z|x))$$



Evidence lower bound

The KL divergence is intractable because it involves $p(x)$, so we optimize the evidence lower bound (ELBO):

$$\mathcal{L}(q) = E[\log p(z, x)] - E[\log q(z)]$$

Mean field approximation

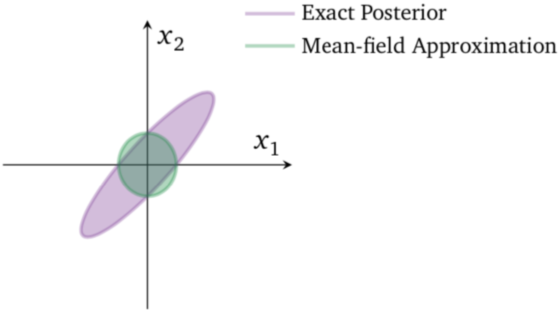
Common simplifying assumption:

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

Leads to closed form updates.

Limitations

Strong assumption of fully factorized posterior fails to capture more complicated posterior distributions.



Blei, Kucukelbir, and McAuliffe (2017)

Expanding the variational family

There has been much work on expanding the variational family $\mathcal{Q}$.

Some examples:

1. Jaakkola and Jordan (1998): Mixture families
2. Miller, Foti, and Adams (2017): Boosting–iteratively adding mixture components
3. Tran, Blei, and Airoldi (2015): Copula families
4. Rezende and Mohamed (2015): Normalizing flows: apply multiple invertible transformations to simple density

Hierarchical variational families

Ranganath, Tran, and Blei (2016)

Impose a model for the variational family:

$$q(z, \lambda) = q(z|\lambda)q(\lambda)$$

Then the marginal posterior approximation for z

$$q(z) = \int q(z|\lambda)q(\lambda)d\lambda$$

can be very complicated as it is an infinite mixture.

Implicit families

Pass noise z (e.g. $\sim N(0, 1)$) through a neural network $G(z)$, then samples from $G(z)$ will follow a complicated distribution.

We can only sample from $G(z)$, cannot evaluate its density, hence an implicit distribution.

This is how the "generator" works in generative adversarial networks.

Implicit families

Tran, Ranganath, and Blei (2017), Huszár (2017), Shi, Sun, and Zhu (2017) use an implicit distribution as $q(z)$.

However estimation becomes difficult. Must estimate density ratio

$$\log \frac{p(x,z)}{q(x,z)}$$

Semi-implicit variational inference (SIVI)

Yin and Zhou (2018)

Combine hierarchical and implicit variational families

$q(z|\psi)$ explicit density (e.g. normal)
$q(\psi)$ implicit density, noise through neural network

Marginal for z

$$q(z) = \int q(z|\psi)q(\psi)d\psi$$

is has very expressive mixture weights, so it too is very expressive.

SIVI Optimization

Yin and Zhou (2018) derive a lower bound for the ELBO under SIVI family.

This can be optimized using Monte Carlo estimates of the gradient, via the reparameterization trick (Kingma and Welling (2013)).

Titsias and Ruiz (2018) show that you can directly compute Monte Carlo estimates of the ELBO under the SIVI family.

Negative binomial example

From Yin and Zhou (2018):

Model

$$\begin{aligned} x_i &\sim NB(r, p) \\ r &\sim Gamma(a, 1/b) \\ p &\sim Beta(\alpha, \beta) \end{aligned}$$

Variational approximation

$$\begin{aligned} q(r, p|\psi) &= LogNormal(r; \mu_r, \sigma_0^2) LogitNormal(p; \mu_p, \sigma_0^2) \\ \psi &= (\mu_r, \mu_p) \sim q(\psi) \\ \sigma_0 &= 0.1 \end{aligned}$$

Negative binomial example

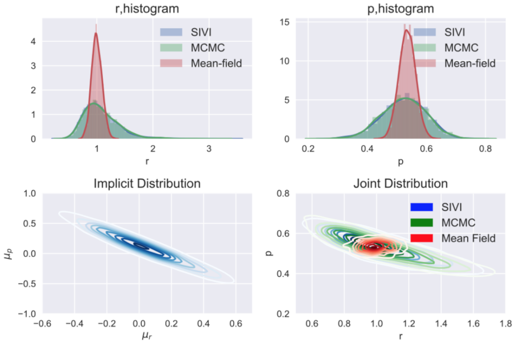


Figure 2. Top row: the marginal posteriors of r and p inferred by MFVI, SIVI, and MCMC. Bottom row: the inferred implicit mixing distribution $q(\psi)$ and joint posterior of r and p.

Yin and Zhou (2018)

Spatial GP regression

In progress...

$$q(\beta, \sigma^2, \tau^2, \phi, \nu) = N(\beta; \mu_\beta, \sigma^2_\beta) \times IG(\sigma^2; a_{\sigma^2}, b_{\sigma^2}) \times$$
$$IG(\tau^2; a_{\tau^2}, b_{\tau^2}) \times U(\phi; a_\phi, b_\phi) \times U(\nu; a_\nu, b_\nu)$$
$$(\mu_\beta, \sigma^2_\beta, a_{\sigma^2}, b_{\sigma^2}, a_{\tau^2}, b_{\tau^2}, a_\phi, b_\phi, a_\nu, b_\nu) \sim q(\psi)$$

Planning on similar comparison to MCMC and mean field VI.

References

Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe. 2017. "Variational Inference: A Review for Statisticians." *Journal of the American Statistical Association* 112 (518): 859–77. https://doi.org/10.1080/01621459.2017.1285773.

Huszár, Ferenc. 2017. "Variational Inference Using Implicit Distributions." *arXiv:1702.08235 [Cs, Stat]*, February. http://arxiv.org/abs/1702.08235.

Jaakkola, Tommi S., and Michael I. Jordan. 1998. *Improving the Mean Field Approximation via the Use of Mixture Distributions*.

Kingma, Diederik P., and Max Welling. 2013. "Auto-Encoding Variational Bayes." *arXiv:1312.6114 [Cs, Stat]*, December. http://arxiv.org/abs/1312.6114.

Miller, Andrew C., Nicholas J. Foti, and Ryan P. Adams. 2017. "Variational Boosting: Iteratively Refining Posterior Approximations." In *International Conference on Machine Learning*, 2420–9. http://proceedings.mlr.press/v70/miller17a.html.

Ranganath, Rajesh, Dustin Tran, and David Blei. 2016. "Hierarchical Variational Models." In *International Conference on Machine Learning*, 324–33. http://proceedings.mlr.press/v48/ranganath16.html.

Rezende, Danilo Jimenez, and Shakir Mohamed. 2015. "Variational Inference with Normalizing Flows." In *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, 1530–8. ICML'15. Lille, France: JMLR.org. http://dl.acm.org/citation.cfm?id=3045118.3045281.

Shi, Jiaxin, Shengyang Sun, and Jun Zhu. 2017. "Kernel Implicit Variational Inference," May. https://arxiv.org/abs/1705.10119v3.

Titsias, Michalis K., and Francisco J. R. Ruiz. 2018. "Unbiased Implicit Variational Inference," August. https://arxiv.org/abs/1808.02078v3.

Tran, Dustin, David Blei, and Edo M Airoldi. 2015. "Copula Variational Inference." In *Advances in Neural Information Processing Systems 28*, edited by C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, 3564–72. Curran Associates, Inc. http://papers.nips.cc/paper/5669-copula-variational-inference.pdf.

Tran, Dustin, Rajesh Ranganath, and David Blei. 2017. "Hierarchical Implicit Models and Likelihood-Free Variational Inference." In *Advances in Neural Information Processing Systems 30*, edited by I. Guyon, U. V.